RAPID DIRECT SEQUENCE ANALYSIS OF MULTI-EXON GENES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims benefit of U.S. Provisional Application No. 60/433,774, filed December 17, 2002. Application Serial No. 60/433,774, filed December 17, 2002, is hereby incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

The research described herein was supported by the Parent Project Muscular Dystrophy, the Muscular Dystrophy Association, the Primary Children's Research Foundation and the National Institutes of Health (NIH R01 NS43264-01 and NIH U01 HG02138-04). The U.S. Government has certain rights in this invention.

FIELD

The compositions, materials, methods, and devices disclosed herein relate to a Single Condition Amplification/Internal Primer (SCAIP) sequencing method for direct sequence analysis of large multi-exon genes from genomic DNA samples and identifying mutations in multi-exon genes. Also, disclosed are methods for diagnosing dystrophinopathies in patients. The disclosed compositions, materials, methods, and devices further relate to compositions for PCR primer sets and sequencing primer sets recognizing the exons or proximal promoter regions for the dystrophin gene.

BACKGROUND

The dystrophinopathies, Duchenne Muscular Dystrophy (DMD) and Becker Muscular Dystrophy (BMD), are the most common inherited disorders of muscle. The prevalence of DMD is generally estimated at 1:3500 live male births (Emery (1991) Neuromuscul Disord 1:19-29). The dystrophin gene is located at Xp21 and is comprised of 79 exons and 8 tissue-specific promoters distributed across approximately 2.2 million base pairs of genomic sequence, making dystrophin the largest gene yet described. Both DMD and BMD are due to mutations in the dystrophin gene. Dystrophin gene deletions are found in approximately 55% of Becker and 65% of Duchenne patients; point mutations account for around 30% of mutations and duplications account for the remainder (Miller *et al.* (1994) Neurol Clin 12:699-725).

Genetic testing for deletions has relied upon a multiplex PCR technique with amplification of fragments containing 18 to 25 of the 79 exons for the gene (Beggs *et al.* (1990) Hum Genet 86:45-48; Chamberlain *et al.* (1990) Multiplex PCR for the diagnosis of Duchenne muscular dystrophy. In: Innis *et al.* (eds) PCR Protocols: A Guide to Methods

and Applications. Academic Press, San Francisco, pp. 272-281) and deletions detected as absent or size-shifted bands on agarose gel analysis. Deletions tend to occur in "hotspots" within the dystrophin gene, and it is estimated that 98% of all dystrophin deletions are detectable by this method.

Testing for dystrophin point mutations has only been available on a research basis from specialized laboratories. Such analysis requires sequencing of all 79 exons and eight promoters. There are no particularly common point mutations or point mutation hotspots currently known, and each affected family may carry a unique mutation in this enormous gene (so-called "private mutations" as they are exclusive to individual families). Instead of direct sequence analysis, some research laboratories perform point mutation analysis on cDNA derived by reverse transcription-PCR (RT-PCR) from muscle mRNA. As an alternative, other laboratories have utilized the protein truncation test (PTT), which may be performed using peripheral blood lymphocyte DNA (Roest *et al.* (1993) Neuromuscul Disord 3:391-394) but often uses mRNA derived from muscle biopsy (Tuffery-Giraud *et al.* (1999) Hum Mutat 14:359-368). There is a drawback to approaches that require muscle biopsy, an invasive procedure with a generally accepted risk of complications (bleeding, infections, hematoma formation) of around 1%, and one that may often be associated with psychological distress for children.

Direct sequence analysis of the dystrophin gene has been considered too labor-intensive, expensive, and time-consuming (Bennett *et al.* (2001) BMC Genet 2:17), but several groups have recently developed strategies to detect exonic sequence variations by screening methods, followed by direct sequence analysis of only variant fragments. One of these strategies is based on single-strand conformational polymorphism (SSCP) analysis (Mendell *et al.* (2001) Neurology 57:645-650). This strategy relies on multiplexing up to 23 amplicons per lane with SSCP in up to five conditions. Mendell *et al.* report that up to 75% of non-deletion mutations may be detected by this method, but there are several drawbacks. One is that all band variations detected by SSCP techniques still need to be sequenced to determine whether they represent pathogenic mutations; the dystrophin gene, because of its size, has many reported polymorphisms. Another problem is that for economies of scale in reagents and technician time, individual samples may need to be saved until multiple samples are available for simultaneous analysis of band variation.

A second screening method relies upon denaturing high-performance liquid chromatography (DHPLC) (Bennett *et al.* (2001) BMC Genet 2:17). This strategy screens

for DNA variations by separating heteroduplex and homoduplex DNA fragments by reverse phase liquid chromatography followed by direct sequence analysis of variant amplicons. Using this method, Bennett *et al.* detected point mutations in 6/8 DNA samples from patients without deletions, and argued for its use on an economic as well as scientific basis (Bennett *et al.* (2001) BMC Genet 2:17). Another screening strategy includes double gradient, denaturing gradient gel electrophoresis (DGGE) (Cremonesi *et al.* (1997) Biotechniques 22:326-330). A drawback to each of these prior art screening methods is the lack of sensitivity. While each method can detect both mutations and non-disease-associated polymorphisms, an additional sequencing step is required to distinguish between these possibilities.

Therefore, in light of the difficulties and short-comings with detecting and characterizing mutations in large multi-exon genes, such as the dystrophin gene, there exists a need for rapid, accurate, and economical sequence analysis of such genes. Disclosed herein are compositions, materials, methods, and devices that satisfy this need.

SUMMARY

In accordance with the purposes of the disclosed compositions, materials, methods, and devices, as embodied and broadly described herein, the disclosed subject matter, in one aspect, relates to a Single Condition Amplification/Internal Primer (SCAIP) sequencing method which allows for the rapid, accurate, and economical analysis of any large multi-exon gene.

An additional aspect of this method is to detect genomic mutations in any large, multi-exon gene including the dystrophin gene.

In accomplishing this and other objects, there has been provided, according to one aspect of the disclosed method, a method relying on amplification of a large number of exons at a single set of PCR temperatures with a first set of amplification primers followed by sequencing without optimization of individual amplicon conditions, using a second, internal set of sequencing primers. The SCAIP sequencing method comprises the steps of:

providing a PCR reaction plate wherein the wells of each plate contain genomic DNA;

adding to each of the wells a different set of left and right PCR primers complementary to a single exonic region or proximal promoter segment for a multi-exon gene of interest and performing a PCR reaction at a uniform set of temperatures;

purifying PCR fragments for the single exonic region or the proximal promoter segment from each of the wells, adding the fragments to a well of a cycle sequencing reaction plate to which is added left and/or right internal sequencing primers corresponding to the single exonic regions or the proximal promoter fragments and sequencing at a uniform set of temperatures;

purification of sequencing products followed by electrophoretic separation and fluorescent detection of nucleotides on a sequence analyzer; and

nucleotide sequence characterization.

More generally, some forms of the disclosed methods involve amplification of a large number of amplicons from a gene or nucleic acid region of interest under the same reaction conditions with a first set of amplification primers followed by sequencing under the same reaction conditions using a second, internal set of sequencing primers. The amplification reactions are preferable carried out simultaneously and/or on the same solid support. The sequencing reactions can be carried out simultaneously and/or on the same solid support. The amplification and sequencing reactions can be carried out on the same solid support (for example, without transfer of amplification products to a different solid support or to different reaction chambers) or different solid supports. Purification of the amplification products prior to sequencing is preferred but not required. The general method can comprise the steps of:

adding to each of a plurality of reaction chambers a nucleic acid sample and a different set of amplification primers, wherein each set of amplification primers is complementary to a single amplicon segment of a gene or nucleic acid region of interest (such as an exonic region or proximal promoter segment of a multi-exon gene of interest) and performing an amplification reaction for each reaction chamber under the same reaction conditions;

bringing into contact in each of a plurality of reaction chambers an amplicon from a different one of the amplification reactions and one or more sequencing primers corresponding to the amplicon and performing a sequencing reaction for each reaction chamber under the same reaction conditions; and

analyzing the sequences of the amplicons.

The nucleic acid sample generally will be the same for each of the reaction chambers in a set of reactions for the analysis of a gene or nucleic acid region of interest. Each reaction chamber is used to amplify and/or sequence a different amplicon from the

gene or nucleic acid region of interest. Useful forms of the method involve amplifying and sequencing all relevant amplicons in the gene or nucleic acid region of interest.

Pursuant to another aspect, the disclosed methods provide for a method of diagnosing mutations in a large multi-exon gene. Individuals may also be tested using the method to identify their status as carriers of DMD or BMD.

Another aspect of the disclosed methods and compositions is the specific amplifying and sequencing primers for the dystrophin gene and their use in a detection kit for DMD or BMD mutations.

Additional advantages of the disclosed methods and compositions will be set forth in part in the description which follows, and in part will be understood from the description, or may be learned by practice of the disclosed method and compositions. The advantages of the disclosed method and compositions will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate several embodiments of the disclosed method and compositions and together with the description, serve to explain the principles of the disclosed method and compositions.

Figure 1 is an agarose gel analysis of primary PCR products from a multi-exon deletion case missing exons 20 to 30 and the DMD260 promoter.

Figure 2 is a graph of the average Phrap score coverage of DMD exons and promoter regions.

DETAILED DESCRIPTION

The compositions, materials, methods, and devices described herein may be understood more readily by reference to the following detailed description of specific aspects of the disclosed subject matter, and methods and the Examples included therein and to the Figures and their previous and following description.

Also, throughout this specification, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this pertains.

The references disclosed are also individually and specifically incorporated by reference herein for the material contained in them that is discussed in the sentence in which the reference is relied upon.

Before the present compositions, materials, methods, and devices, are disclosed and described, it is to be understood that the aspects described below are not limited to specific synthetic methods or specific reagents, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular aspects only and is not intended to be limiting.

Disclosed herein are materials, compositions, and components that can be used for, can be used in conjunction with, can be used in preparation for, or are products of the disclosed method and compositions. These and other materials are disclosed herein, and it is understood that when combinations, subsets, interactions, groups, etc. of these materials are disclosed that while specific reference of each various individual and collective combinations and permutation of these compounds may not be explicitly disclosed, each is specifically contemplated and described herein. For example, if an internal primer is disclosed and discussed and a number of modifications that can be made to a number of molecules including the internal primer are discussed, each and every combination and permutation of the internal primer and the modifications that are possible are specifically contemplated unless specifically indicated to the contrary. Thus, if a class of molecules A, B, and C are disclosed as well as a class of molecules D, E, and F and an example of a combination molecule, A-D is disclosed, then even if each is not individually recited, each is individually and collectively contemplated. Thus, is this example, each of the combinations A-E, A-F, B-D, B-E, B-F, C-D, C-E, and C-F are specifically contemplated and should be considered disclosed from disclosure of A, B, and C; D, E, and F; and the example combination A-D. Likewise, any subset or combination of these is also specifically contemplated and disclosed. Thus, for example, the sub-group of A-E, B-F, and C-E are specifically contemplated and should be considered disclosed from disclosure of A, B, and C; D, E, and F; and the example combination A-D. This concept applies to all aspects of this application including, but not limited to, steps in methods of making and using the disclosed compositions. Thus, if there are a variety of additional steps that can be performed it is understood that each of these additional steps can be performed with any specific embodiment or combination of embodiments of the disclosed methods, and that each such combination is specifically contemplated and should be considered disclosed.

A. General Definitions:

In this specification and in the claims that follow, reference will be made to a number of terms, which shall be defined to have the following meanings:

As used in the specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a nucleotide" includes mixtures of two or more such nucleotides, reference to "an amino acid" includes mixtures of two or more such amino acids, reference to "the primer" includes mixtures of two or more such primers, and the like.

"Optional" or "optionally" means that the subsequently described event or circumstance can or cannot occur, and that the description includes instances where the event or circumstance occurs and instances where it does not. For example, the phrase "amplicons can optionally be purified" means that the amplicons may or may not be purified and that the description includes both methods where the amplicons are purified and methods where the amplicons are not purified.

Ranges may be expressed herein as from "about" one particular value, and/or to "about" another particular value. When such a range is expressed, another aspect includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms another aspect. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint.

"Individual," as used herein, means a subject. In one aspect, the individual is a mammal such as a primate, and, in another aspect, the individual is a human. The term "individual" also includes domesticated animals (*e.g.*, cats, dogs, etc.), livestock (*e.g.*, cattle, horses, pigs, sheep, goats, etc.), and laboratory animals (*e.g.*, mouse, rabbit, rat, guinea pig, etc.).

There are a variety of molecules disclosed herein that are nucleic acid based, including for example the nucleic acids that encode, for example, dystrophin as well as any other proteins disclosed herein, as well as various functional nucleic acids. The disclosed nucleic acids are made up of for example, nucleotides, nucleotide analogs, or nucleotide substitutes. Non-limiting examples of these and other molecules are discussed herein.

A nucleotide is a molecule that contains a base moiety, a sugar moiety and a phosphate moiety. Nucleotides can be linked together through their phosphate moieties

and sugar moieties creating an internucleoside linkage. The base moiety of a nucleotide can be adenin-9-yl (A), cytosin-1-yl (C), guanin-9-yl (G), uracil-1-yl (U), and thymin-1-yl (T). The sugar moiety of a nucleotide is a ribose or a deoxyribose. The phosphate moiety of a nucleotide is pentavalent phosphate. An non-limiting example of a nucleotide would be 3'-AMP (3'-adenosine monophosphate) or 5'-GMP (5'-guanosine monophosphate).

A nucleotide analog is a nucleotide which contains some type of modification to either the base, sugar, or phosphate moieties. Modifications to nucleotides are well known in the art and would include for example, 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, and 2-aminoadenine as well as modifications at the sugar or phosphate moieties.

Nucleotide substitutes are molecules having similar functional properties to nucleotides, but which do not contain a phosphate moiety, such as peptide nucleic acid (PNA). Nucleotide substitutes are molecules that will recognize nucleic acids in a Watson-Crick or Hoogsteen manner, but which are linked together through a moiety other than a phosphate moiety. Nucleotide substitutes are able to conform to a double helix type structure when interacting with the appropriate target nucleic acid.

It is also possible to link other types of molecules (conjugates) to nucleotides or nucleotide analogs to enhance for example, cellular uptake. Conjugates can be chemically linked to the nucleotide or nucleotide analogs. Such conjugates include but are not limited to lipid moieties such as a cholesterol moiety. (Letsinger et al., Proc. Natl. Acad. Sci. USA, 1989,86, 6553-6556).

A Watson-Crick interaction is at least one interaction with the Watson-Crick face of a nucleotide, nucleotide analog, or nucleotide substitute. The Watson-Crick face of a nucleotide, nucleotide analog, or nucleotide substitute includes the C2, N1, and C6 positions of a purine based nucleotide, nucleotide analog, or nucleotide substitute and the C2, N3, C4 positions of a pyrimidine based nucleotide, nucleotide analog, or nucleotide substitute.

A Hoogsteen interaction is the interaction that takes place on the Hoogsteen face of a nucleotide or nucleotide analog, which is exposed in the major groove of duplex DNA. The Hoogsteen face includes the N7 position and reactive groups (NH_2 or O) at the C6 position of purine nucleotides.

There are a variety of sequences related to, for example, the dystrophin gene as well as any other nucleic acids sequences that are disclosed on GenBank, and these sequences

and others are herein incorporated by reference in their entireties as well as for individual subsequences contained therein.

A variety of sequences are provided herein and these and others can be found in GenBank, at www.ncbi.nlm.nih.gov. Those of skill in the art understand how to resolve sequence discrepancies and differences and to adjust the compositions and methods relating to a particular sequence to other related sequences. Primers and/or probes can be designed for any sequence given the information disclosed herein and known in the art.

Disclosed are compositions including primers and probes, which are capable of interacting with the genes disclosed herein. In certain embodiments the primers are used to support DNA amplification reactions. In other embodiments, the primers are used to support sequencing reactions. Typically the primers will be capable of being extended in a sequence specific manner. Extension of a primer in a sequence specific manner includes any methods wherein the sequence and/or composition of the nucleic acid molecule to which the primer is hybridized or otherwise associated directs or influences the composition or sequence of the product produced by the extension of the primer. Extension of the primer in a sequence specific manner therefore includes, but is not limited to, PCR, DNA sequencing, DNA extension, DNA polymerization, RNA transcription, or reverse transcription. Techniques and conditions that amplify the primer in a sequence specific manner are preferred. In certain embodiments the primers are used for the DNA amplification reactions, such as PCR or direct sequencing. It is understood that in certain embodiments the primers can also be extended using non-enzymatic techniques, where for example, the nucleotides or oligonucleotides used to extend the primer are modified such that they will chemically react to extend the primer in a sequence specific manner. Typically the disclosed primers hybridize with the nucleic acid or region of the nucleic acid or they hybridize with the complement of the nucleic acid or complement of a region of the nucleic acid.

B. Method:

Disclosed herein is a Single Condition Amplification/Internal Primer (SCAIP) sequencing method which allows for the rapid, accurate, and economical analysis of any large multi-exon gene. This method is particularly useful for detecting and characterizing mutations in large multi-exon genes such as the dystrophin gene. Mutations in the dystrophin gene result in both Duchenne and Becker muscular dystrophy (DMD and BMD), as well as X-linked dilated cardiomyopathy. Mutational analysis is complicated by the

large size of the gene, which consists of 79 exons and 8 promoters spread over 2.2 million base pairs of genomic DNA. Deletions of one or more exons account for 55-65% of cases of DMD and BMD. A multiplex PCR method is currently the most widely available method for mutational analysis and it detects approximately 98% of deletions. However, detection of point mutations and small subexonic rearrangements has remained challenging. The disclosed method overcomes the problems associated with prior art DNA screening methods by allowing direct sequence analysis of a multi-exon gene in a rapid, accurate, and economical fashion.

The disclosed method provides for the identification and analysis of specific individual genomic mutations such as deletions, point mutations, frameshifts, or combinations thereof, in gene complexes with multiple exons/introns spanning large genomic regions.

As used herein, the term "deletion" refers to those genomic DNA sequences in which one or more nucleic acid bases has been deleted from the sequence and is no longer present in the gene.

As used herein, the term "point mutation" refers to a mutation resulting from a change in a single base pair in the DNA molecules, caused by the substitution of one nucleotide for another.

As used herein, the term "frameshift" refers to a loss or gain of some number of nucleotides which is not divisible by three (*i.e.*, one or more codons).

The primary determinant of sequence specificity and base call quality is the uniform use of internal sequencing primers. The disclosed assay design is robust in that it can tolerate secondary, non-specific PCR amplification products, as opposed to assays that use a single set of primers or use secondary primers to universal sequences on the 5' end of the PCR primers. An object of the method is the optimization a single 96 well plate assay in which all coding regions and promoters of the dystrophin gene are amplified in a single PCR plate. The PCR products are then purified in plate format using multi-channel pipetting robots, and two cycle sequencing plates prepared and processed. Sequencing can be routinely performed within 3 working days following DNA purification at a reasonable cost including both reagents and personnel costs. The one patient-one plate assay is designed for the requirements of both a rapid turnaround time for the assay, as well as making the assay scalable with a potential increase in demand.

Thus, an embodiment for the methods and compositions disclosed herein is a method designed to achieve PCR amplification and cycle sequencing of 96 distinct amplicons from a single individual using uniform thermal cycling parameters in a single vessel such as a 96 or 384 well thermal cycler microtiter plate. Alternatively, several individuals with multiple amplicons can be assayed in the same plate, *e.g.*, four individuals with twenty-four distinct amplicons. The method comprises: designing PCR and sequence primers with software, performing a PCR reaction with the PCR primers on a DNA sample, performing a sequencing reaction with sequencing primers on the PCR products, electrophoretic separation and fluorescent detection of the sequencing reaction products on a capillary sequencer, and analyzing the DNA sequence with software.

In one aspect, disclosed herein is a method for characterizing the mutations in a multi-exon gene comprising: providing a sample of a patient's purified genomic DNA, plating the DNA in a 96 well plate followed by PCR amplification of gene-specific DNA fragments with a different PCR amplification primer set for each of the 96 wells under uniform amplification conditions. This is followed by cycle sequencing of the amplified DNA fragments with a different internal sequencing primer set for each well in a 96 well plate under uniform sequencing conditions. Samples from each sequencing reaction are then loaded onto an automated DNA capillary sequencer. Sequence data are then collected and analyzed with a computer using a mutation detection software program. A database is generated from the mutation sequence information, and with the software, the product sequence can be compared to other known sequences.

C. Genes:

The disclosed methods can involve the use of any genomic DNA sequence or any other nucleic acid sequence of interest. For example, a genomic DNA sequence to be detected herein can be derived from an organism, preferably a human patient and more preferably a human patient having or suspected of having a dystrophinopathy. The source of the genomic DNA from the organism to be tested can be from any tissue, such as peripheral lymphocytes.

The disclosed method is applicable to known or unknown genes, and should allow the development of widely-available assays for any number of large, multi-exon genes. Examples of some multi-exon genes which are candidates for the use of the disclosed method are NF-1, ATM, dysferlin, calpain, αβγδε sarcoglycans, collagens 6A1-3, Nebulin, and Titin. More preferred are those polymorphic genes associated with orphan diseases

including but not limited to the dystrophin gene in DMD or BMD, the SOD-1 gene in Amyotrophic Lateral Sclerosis, NF-1 in von Recklinghausen neurofibromatosis, and dysferlin in limb-girdle muscular dystrophy type 2B.

D. Amplicons:

For the purposes of the disclosed methods, distinct regions of the nucleic acid sequence of interest, such as a sample of genomic DNA, can be identified for amplification. These regions of the nucleic acid of interest can each be amplified with a set of amplification primers. As such, these distinct regions of a nucleic acid sequence of interest can be termed amplicons. Also, as used herein, the term amplicon refers to the product of an amplification reaction upon a distinct region of a nucleic acid region of interest. Amplicons from a given nucleic acid sequence of interests or genomic DNA can be non-overlapping regions of the nucleic acid sequence of interest. Alternatively, amplicons can have overlapping portions in the nucleic acid sequence of interest. Also, an amplicon can be, for example, a single exon, a single exonic region or a proximal promoter sequence.

An amplicon can be of any length. For example, a amplicon can have an average length of, 0.5 kilobases (kb), 0.6 kb, 0.7 kb, 0.8 kb, 0.9 kb, 1.0 kb, 1.1 kb, 1.2 kb, 1.3 kb, 1.4 kb, 1.5 kb, 1.6 kb, 1.7 kb, 1.8 kb, 1.9 kb, 2.0 kb, 2.2 kb, 2.5 kb, 3 kb, 3.5 kb, 4 kb, 4.5 kb, 5 kb, 5.5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 11 kb, 12 kb, 13 kb, 14 kb, 15 kb, 16 kb, 18 kb, 20 kb, 22 kb, 24 kb, 26 kb, 28 kb, 30 kb, 2 kb or more, 2.5 kb or more, 3 kb or more, 3.5 kb or more, 4 kb or more, 4.5 kb or more, 5 kb or more, 5.5 kb or more, 6 kb or more, 7 kb or more, 8 kb or more, 9 kb or more, 10 kb or more, 11 kb or more, 12 kb or more, 13 kb or more, 14 kb or more, 15 kb or more, 16 kb or more, 18 kb or more, 20 kb or more, 22 kb or more, 24 kb or more, 26 kb or more, 28 kb or more, 30 kb or more, about 2 kb, about 2.5 kb, about 3 kb, about 3.5 kb, about 4 kb, about 4.5 kb, about 5 kb, about 5.5 kb, about 6 kb, about 7 kb, about 8 kb, about 9 kb, about 10 kb, about 11 kb, about 12 kb, about 13 kb, about 14 kb, about 15 kb, about 16 kb, about 18 kb, about 20 kb, about 22 kb, about 24 kb, about 26 kb, about 28 kb, about 30 kb, about 2 kb or more, about 2.5 kb or more, about 3 kb or more, about 3.5 kb or more, about 4 kb or more, about 4.5 kb or more, about 5 kb or more, about 5.5 kb or more, about 6 kb or more, about 7 kb or more, about 8 kb or more, about 9 kb or more, about 10 kb or more, about 11 kb or more, about 12 kb or more, about 13 kb or more, about 14 kb or more, about 15 kb or more, about 16 kb or more, about 18 kb or more, about 20 kb or more, about 22 kb or more, about 24 kb or more, about 26 kb or more, about 28 kb or more, or about 30 kb or more. In some aspects, the amplicon has an

average length of from about 1.0 kb to about 2.0 kb, from about 1.0 kb to about 1.8 kb, from about 1.0 kb to about 1.6 kb, from about 1.0 kb to about 1.4 kb, from about 1.0 kb to about 1.2 kb, from about 1.2 kb, to about 2.0 kb, from about 1.2 kb to about 1.8 kb, from about 1.2 kb to about 1.6 kb, from about 1.2 kb to about 1.4 kb, from about 1.4 kb to about 2.0 kb, from about 1.8 kb, from about 1.4 kb to about 1.6 kb, from about 1.6 kb to about 2.0 kb, from about 1.6 kb to about 1.8 kb, or from about 1.8 kb to about 2.0 kb. In another aspect, the amplicon can have an average length of from about 1.2 to about 1.4 kb.

While amplicons can be of any length (as measured by the number of nucleotides in the amplicon), it is useful to note that having larger amplicons will require fewer reaction chambers when practicing the methods disclosed herein. Conversely, the smaller the amplicon size, the more reaction chambers that are needed. For example, partitioning a nucleic acid sequence of interest into, say, 50 amplicons, will require more reaction chambers than it would if the nucleic acid sequence were partitioned into, say, 25 amplicons.

Also, there is no specific requirement that a certain number of amplicons be used in the methods disclosed herein. The number of amplicons will largely depend on the size of the nucleic acid sequence of interest or genomic DNA. In general, a large nucleic acid sequences of interest will typically result in a larger number of amplicons. Similarly, smaller nucleic acid sequences will typically result in less amplicons being used. However, in the disclosed methods, any number of amplicons can be used. In one aspect, the number of amplicons that can be used in the methods disclosed herein are about 48, about 96, or about 348. In another aspect, the number of amplicons that can be used are, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229,

230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, or 348 amplicons. It is also, possible to perform the disclosed method on more than 348 amplicons, such as about 350, 400, 450, 500, 600, 750, 1000, 1250, 1500, 2000, 2500, 3000, 4000, or 5000 amplicons.

Also, according to the disclosed methods, a plurality of amplicons are amplified in a plurality of reaction chambers. It is useful for such amplification reactions to be conducted at similar or the same conditions. To this end, it can be beneficial to have amplicons of substantially similar lengths. In this way, the amplification conditions for each amplicon will be similar, and the amplification of more than one amplicon will be more efficient. For example, amplicons of similar lengths can be amplified to a similar extent at substantially the same temperature, with substantially the same amount of reagents, and with the same number of cycles.

E. Reaction Chambers:

The disclosed methods, either in whole or in part, can be performed in or on solid supports or in or on reaction chambers. For example, the disclosed amplification and sequencing steps (or any other operations of the disclosed methods) can be performed with the reaction mixture in or on solid supports or in or on reaction chambers. For example, the disclosed amplification and sequencing can be performed with the reaction mixture on solid supports having reaction chambers. A reaction chamber is any structure in which a separate reaction can be performed. Useful reaction chambers include tubes, test tubes, eppendorf tubes, vessels, micro vessels, plates, wells, wells of micro well plates, wells of microtitre plates, chambers, micro fluidics chambers, micro machined chambers, sealed chambers, holes, depressions, dimples, dishes, surfaces, membranes, microarrays, fibers, glass fibers, optical fibers, woven fibers, films, beads, bottles, chips, compact disks, shaped polymers, particles, microparticles or other structures that can support separate reactions. Reaction chambers can be made from any suitable material, such as solid support materials. Such materials include acrylamide, cellulose, nitrocellulose, glass, gold, polystyrene, polyethylene vinyl acetate, polypropylene, polymethacrylate, polyethylene, polyethylene

oxide, glass, polysilicates, polycarbonates, teflon, fluorocarbons, nylon, silicon rubber, polyanhydrides, polyglycolic acid, polylactic acid, polyorthoesters, functionalized silane, polypropylfumerate, collagen, glycosaminoglycans, and polyamino acids. Solid supports preferably comprise arrays of reaction chambers. Solid supports and reaction chambers can be porous or non-porous. A useful form for reaction chambers is a microtiter dish. A particularly useful form of microtiter dish is the standard 96-well type. In some embodiments, a multiwell glass slide can be employed.

In connection with reaction chambers, a separate reaction refers to a reaction where substantially no cross contamination of reactants or products will occur between different reaction chambers. Substantially no cross contamination refers to a level of contamination of reactants or products below a level that would be detected in the particular reaction or assay involved. For example, if nucleic acid contamination from another reaction chamber would not be detected in a given reaction chamber in a given assay (even though it may be present), there is no substantial cross contamination of the nucleic acid. It is understood, therefore, that reaction chambers can comprise, for example, locations on a planar surface, such as spots, so long as the reactions performed at the locations remain separate and are not subject to mixing. Some useful forms of the disclosed methods can use reaction chambers that can be sealed to allow thermocycle reactions (for example, PCR and cycle sequencing) of small volumes.

Methods for immobilization of nucleic acid sequences to solid-state substrates are well established. For example, suitable attachment methods are described by Pease *et al.*, *Proc. Natl. Acad. Sci. USA* **91**(11):5022-5026 (1994), and Khrapko *et al.*, *Mol Biol (Mosk) (USSR)* 25:718-730 (1991). A method for immobilization of 3'-amine oligonucleotides on casein-coated slides is described by Stimpson *et al.*, *Proc. Natl. Acad. Sci. USA* **92**:6379-6383 (1995). A useful method of attaching oligonucleotides to solid-state substrates is described by Guo *et al.*, *Nucleic Acids Res*. 22:5456-5465 (1994).

Components can be associated or immobilized on a solid support at any density. Components can be immobilized to the solid support at a density exceeding 400 different components per cubic centimeter. Arrays of components can have any number of components. For example, an array can have at least 1,000 different components immobilized on the solid support, at least 10,000 different components immobilized on the solid support, at least 100,000 different components immobilized on the solid support, or at least 1,000,000 different components immobilized on the solid support.

In one aspect, the disclosed method can involve simultaneously performing various reactions, such as amplification and sequencing, on a plurality of amplicons. It is preferable that these reactions be conducted on an a plurality of amplicons where each amplicon has been allocated to a separate reaction chamber. That is, one amplicon can amplified and/or sequenced in one reaction chamber. However, although not preferred, more than one amplicon, i.e., 2, 3, 4, 5, 10, 20, etc., can be amplified and/or sequenced in one reaction chamber. Also, the same amplicon can be amplified and/or sequenced in multiple reaction chambers. This could be done, for example, when the additional reaction chambers are used as controls or duplicates. It is preferable that multiple reactions be conducted in or on a single solid support, preferably with a plurality of reaction chambers. That is, multiple amplicon, such as all of the amplicons for a multi-exon gene, can be amplified and/or sequenced on one solid support. However, multiple amplicons for a multi-exon gene can also be amplified and/or sequenced on multiple solid supports.

The disclosed methods can involve the use of multiple reaction chambers. For example, in one aspect, the disclosed methods can involve amplifications reactions that are simultaneously carried out on the contents of various reaction chambers. Similarly, the disclosed methods can involve sequencing reactions that are simultaneously carried out on the contents of various reaction chambers. The number of reaction chambers can be related to the number of amplicons, such as one reaction chamber for each amplicon. While the number of reaction chambers can be the same as the number of amplicons, additional reaction chambers can also be used for controls or duplicates. In one aspect, the disclosed methods can utilize 48, 96, or 348 reaction chambers. In another aspect, the disclosed methods contemplates that 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221,

222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, or 348 reaction chambers are used. It is also, possible to perform the disclose method on more than 348 reaction chambers, such as about 350, 400, 450, 500, 600, 750, 1000, 1250, 1500, 2000, 2500, 3000, 4000, or 5000 reaction chambers.

In one aspect of the disclosed methods, a nucleic acid sample (such as a genomic sample) containing the nucleic acid sequence of interest (such as a multi-exon gene) is contacted with, i.e., placed in or immobilized on, a reaction chamber or solid support before any amplification primers are added. Alternatively, amplification primers can be contacted with the reaction chamber or solid support prior to the introduction of any nucleic acid samples. More generally, components present in the reactions disclosed herein can be mixed, added or combined in any order, in any combination, or simultaneously.

F. Amplification and Sequencing Primers:

Amplification and sequencing reactions can be performed on a plurality of amplicons in a plurality of reaction chambers. As such, these amplification and sequencing reactions utilize sets of amplification primers and sets of sequencing primers. The PCR amplification and sequencing primers are selected to be complementary to the different strands of each specific sequence to be amplified. Primer's can be designed using any known primer prediction software program such as Oligo, GeneFisher, Web Primer or Primer 3 software (a primer prediction program with user-definable parameters for Tm, GC-hairpins, etc.).

For primer prediction of a multi-exon gene, such as dystrophin, dysferlin, calpain, or collagen VI, the genomic sequence is first prepared by masking all known human sequence repeats using the RepeatMasker program. Sequence repeats are re-analyzed when choosing sequence primers and unique repeats are unmasked. The genomic sequence is also masked when choosing sequence primers by a Perl script to eliminate single base repeats (AAAA or GGGG) occurring in the sequence primer. Perl script uses the RNA cross-match output (pair-wise Smith-Waterman comparison) of the mRNA against the genomic sequence to

isolate the exon sequence and flanking genomic sequence. Size parameters passed to the Perl script determine the size of the PCR product. The Perl script generates a Primer 3-formatted sequence file. Primer 3 can generate four potential primer sets, and the primers are cross-matched against the consensus genomic and primer positions relative to the exons. An example of the Perl script is shown in the Program Listing below.

According to the disclosed methods, a set of right and left amplification primers are used for each amplicon. It is preferable that a different set of amplification primers be used for each amplicon. The sequencing primers are preferably internal to the PCR primers, increasing the tolerance to non-specific amplification products in the PCR stage. Just a single sequencing primer can be used. Preferably, however, two sequencing primers are used. The two sequencing primers can be forward and reverse primers or, alternatively, two forward primers or two reverse primers. The use of a forward and reverse internal sequencing primer can relax the stringency needed to get robust amplification of multiple different amplicons under uniform thermal cycling conditions.

Primers for use in the disclosed methods are oligonucleotides having sequence complementary to the target sequence, such as a nucleic acid sequence of interest, an amplicon of a nucleic acid sequence of interest, or an exon or proximal promoter of a nucleic acid sequence of interest. This sequence is referred to as the complementary portion of the primer. The complementary portion of a primer can be any length that supports specific and stable hybridization between the primer and the target sequence under the reaction conditions. Generally, this can be 10 to 35 nucleotides long or 16 to 24 nucleotides long. In some aspects, the primers can be from 5 to 60 nucleotides long, and in particular, can be 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, and/or 20 nucleotides long.

The disclosed amplification and sequence primers can have one or more modified nucleotides. Such primers are referred to herein as modified primers. Modified primers have several advantages. First, some forms of modified primers, such as RNA/ 2'-O-methyl RNA chimeric primers, have a higher melting temperature (Tm) than DNA primers. This increases the stability of primer hybridization and will increase strand invasion by the primers. This will lead to more efficient priming. Also, since the primers are made of RNA, they will be exonuclease resistant. Such primers, if tagged with minor groove binders at their 5' end, will also have better strand invasion of the template dsDNA.

Chimeric primers can also be used. Chimeric primers are primers having at least two types of nucleotides, such as both deoxyribonucleotides and ribonucleotides,

ribonucleotides and modified nucleotides, or two different types of modified nucleotides. One form of chimeric primer is peptide nucleic acid/nucleic acid primers. For example, 5'-PNA-DNA-3' or 5'-PNA-RNA-3' primers may be used for more efficient strand invasion and polymerization invasion. The DNA and RNA portions of such primers can have random or degenerate sequences. Other forms of chimeric primers are, for example, 5'- (2'-O-Methyl) RNA-RNA-3' or 5'- (2'-O-Methyl) RNA-DNA-3'.

Many modified nucleotides (nucleotide analogs) are known and can be used in oligonucleotides. A nucleotide analog is a nucleotide which contains some type of modification to either the base, sugar, or phosphate moieties. Modifications to the base moiety would include natural and synthetic modifications of A, C, G, and T/U as well as different purine or pyrimidine bases, such as uracil-5-yl, hypoxanthin-9-yl (I), and 2-aminoadenin-9-yl. A modified base includes but is not limited to 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Additional base modifications can be found for example in U.S. Pat. No. 3,687,808, Englisch et al., Angewandte Chemie, International Edition, 1991, 30, 613, and Sanghvi, Y. S., Chapter 15, Antisense Research and Applications, pages 289-302, Crooke, S. T. and Lebleu, B. ed., CRC Press, 1993. Certain nucleotide analogs, such as 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine can increase the stability of duplex formation. Other modified bases are those that function as universal bases. Universal bases include 3-nitropyrrole and 5-nitroindole. Universal bases substitute for the normal bases but have no bias in base pairing. That is, universal bases can base pair with any other base. Primers composed, either in whole or in part, of nucleotides with universal bases are useful for reducing or eliminating amplification bias against repeated sequences in a target sample. This would be useful, for example, where a loss of sequence complexity in the amplified products is

undesirable. Base modifications often can be combined with for example a sugar modification, such as 2'-O-methoxyethyl, to achieve unique properties such as increased duplex stability. There are numerous United States patents such as 4,845,205; 5,130,302; 5,134,066; 5,175,273; 5,367,066; 5,432,272; 5,457,187; 5,459,255; 5,484,908; 5,502,177; 5,525,711; 5,552,540; 5,587,469; 5,594,121, 5,596,091; 5,614,617; and 5,681,941, which detail and describe a range of base modifications. Each of these patents is herein incorporated by reference.

Nucleotide analogs can also include modifications of the sugar moiety. Modifications to the sugar moiety would include natural modifications of the ribose and deoxyribose as well as synthetic modifications. Sugar modifications include but are not limited to the following modifications at the 2' position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C1 to C10, alkyl or C2 to C10 alkenyl and alkynyl. 2' sugar modifications also include but are not limited to $-O[(CH_2)n\ O]m\ CH_3$, $-O(CH_2)n\ OCH_3$, $-O(CH_2)n\ NH_2$, $-O(CH_2)n\ CH_3$, $-O(CH_2)n\ -ONH_2$, and $-O(CH_2)nON[(CH_2)n\ CH_3)]_2$, where n and m are from 1 to about 10.

Other modifications at the 2' position include but are not limited to: C1 to C10 lower alkyl, substituted lower alkyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, $SO_2\ CH_3$, ONO_2, NO_2, N_3, NH_2, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. Similar modifications may also be made at other positions on the sugar, particularly the 3' position of the sugar on the 3' terminal nucleotide or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Modified sugars would also include those that contain modifications at the bridging ring oxygen, such as CH_2 and S. Nucleotide sugar analogs may also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar. There are numerous United States patents that teach the preparation of such modified sugar structures such as 4,981,957; 5,118,800; 5,319,080; 5,359,044; 5,393,878; 5,446,137; 5,466,786; 5,514,785; 5,519,134; 5,567,811; 5,576,427; 5,591,722; 5,597,909; 5,610,300; 5,627,053; 5,639,873; 5,646,265; 5,658,873; 5,670,633; and 5,700,920, each of which is herein incorporated by reference in its entirety.

Nucleotide analogs can also be modified at the phosphate moiety. Modified phosphate moieties include but are not limited to those that can be modified so that the linkage between two nucleotides contains a phosphorothioate, chiral phosphorothioate, phosphorodithioate, phosphotriester, aminoalkylphosphotriester, methyl and other alkyl phosphonates including 3'-alkylene phosphonate and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates. It is understood that these phosphate or modified phosphate linkages between two nucleotides can be through a 3'-5' linkage or a 2'-5' linkage, and the linkage can contain inverted polarity such as 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms are also included. Numerous United States patents teach how to make and use nucleotides containing modified phosphates and include but are not limited to, 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, each of which is herein incorporated by reference.

It is understood that nucleotide analogs need only contain a single modification, but may also contain multiple modifications within one of the moieties or between different moieties.

Nucleotide substitutes are molecules having similar functional properties to nucleotides, but which do not contain a phosphate moiety, such as peptide nucleic acid (PNA). Nucleotide substitutes are molecules that will recognize and hybridize to complementary nucleic acids in a Watson-Crick or Hoogsteen manner, but which are linked together through a moiety other than a phosphate moiety. Nucleotide substitutes are able to conform to a double helix type structure when interacting with the appropriate target nucleic acid.

Nucleotide substitutes are nucleotides or nucleotide analogs that have had the phosphate moiety and/or sugar moieties replaced. Nucleotide substitutes do not contain a standard phosphorus atom. Substitutes for the phosphate can be for example, short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone

backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH2 component parts. Numerous United States patents disclose how to make and use these types of phosphate replacements and include but are not limited to 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437; and 5,677,439, each of which is herein incorporated by reference.

It is also understood in a nucleotide substitute that both the sugar and the phosphate moieties of the nucleotide can be replaced, by for example an amide type linkage (aminoethylglycine) (PNA). United States patents 5,539,082; 5,714,331; and 5,719,262 teach how to make and use PNA molecules, each of which is herein incorporated by reference. (See also Nielsen *et al.*, *Science* 254:1497-1500 (1991)).

Primers can be comprised of nucleotides and can be made up of different types of nucleotides or the same type of nucleotides. For example, one or more of the nucleotides in a primer can be ribonucleotides, 2'-O-methyl ribonucleotides, or a mixture of ribonucleotides and 2'-O-methyl ribonucleotides; about 10% to about 50% of the nucleotides can be ribonucleotides, 2'-O-methyl ribonucleotides, or a mixture of ribonucleotides and 2'-O-methyl ribonucleotides; about 50% or more of the nucleotides can be ribonucleotides, 2'-O-methyl ribonucleotides, or a mixture of ribonucleotides and 2'-O-methyl ribonucleotides; or all of the nucleotides are ribonucleotides, 2'-O-methyl ribonucleotides, or a mixture of ribonucleotides and 2'-O-methyl ribonucleotides. The nucleotides can be comprised of bases (that is, the base portion of the nucleotide) and can (and normally will) comprise different types of bases. For example, one or more of the bases can be universal bases, such as 3-nitropyrrole or 5-nitroindole; about 10% to about 50% of the bases can be universal bases; about 50% or more of the bases can be universal bases; or all of the bases can be universal bases.

A particularly useful embodiment of the disclosed methods is a method for detecting mutations in the dystrophin gene. The disclosed method is at least as sensitive as DOVAM screening, and has been successful in identifing at least one mutation undetected by the DOVAM method. Sequencing specificity is gained by uniform use of a second, internal set

of sequencing primers. Sufficient sequencing specificity is obtained without optimization of individual amplicon conditions. The disclosed method results in complete double-stranded sequencing coverage of all known coding regions and 7 of the 8 tissue-specific promoters. Although the dystrophin muscle isoform coding region consists of 11.1 kb, the disclosed sequencing method analyzes an average of nearly 110 kb of sequence, allowing detection of polymorphisms in flanking intronic regions as well as the 3' UTR and 5' regions. The disclosed method allows detection of the approximately 2% of patients with exonic deletions not detected by the widely available multiplex PCR technique. The disclosed method gives highly reproducible and accurate results, and can be performed economically on single samples as described in further detail hereinafter.

The amplification and/or sequence primers can be any size that supports the desired enzymatic manipulation of the primer, such as amplification and/or sequencing. A typical primer would be at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more nucleotides long.

G. PCR:

Various thermocycling parameters and PCR enzyme/buffer combinations that are known in the art may be used to arrive at a single condition for amplification of DNA fragments (Maniatis, T., E. F. Fritsch and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual). After the PCR reaction is complete, the amplification products from each reaction chamber can optionally be purified. Purification techniques are known in the art. The examples below illustrate techniques for such purification. The purified or unpurified amplification products from each reaction chamber can be transferred to a second reaction chamber. Alternatively, the purified or unpurified amplification products can be left in the same reaction chamber.

H. Sequencing:

According to the disclosed methods, the amplicons can be sequenced under uniform temperature and conditions. The internal sequencing primers are added to a reaction chamber. This reaction chamber may be the same reaction chamber used in the PCR amplification, and will thus contain the purified or unpurified amplified amplicons. Alternatively, the internal sequencing primers can be added to a second reaction chamber

prior to, during, or after amplified amplicons have been transferred from the original reaction chamber used in the amplification reaction.

The disclosed method is adaptable for any sequencing method or detection method that relies upon or includes chain extension. These methods include, but are not limited to, sequencing methods based upon Sanger sequencing, and detection methods, such as primer oligo base extension (PROBE) (see, e.g., U.S. Pat. No. 6,043,031 and U.S. Pat. No. 6,235,478), that include a step of chain extension. Automated techniques have also been developed to increase the throughput and decrease the cost of nucleic acid sequencing methods, e.g., U.S. Pat. No. 5,171,534; Connell et al., Biotechniques, 5(4): 342-348 (1987); and Trainor, Anal. Chem., 62: 418-426 (1990). Numerous useful sequencing techniques, including, for example, cycle sequencing, are known and can be adapted for use in the disclosed method.

I. Kits:

The materials described above as well as other materials can be packaged together in any suitable combination as a kit useful for performing, or aiding in the performance of, the disclosed method. It is useful if the kit components in a given kit are designed and adapted for use together in the disclosed method. For example disclosed are kits for the detection and, optionally, characterization, of mutations in multi-exon genes, the kit comprising sets of amplification primers and sets of internal sequencing primers that are designed for the particular multi-exon gene. The kits also can contain reaction chambers or solid supports, amplicons from the multi-exon gene, amplification and/or sequencing reagents, solvents, probes, markers, detection tags, and the like. Also disclosed are kits for the detection and, optionally, characterization, of mutations in the dystrophin gene, the kit comprising sets of amplification primers and sets of internal sequence primers. The kits can also contain amplicons from the dystrophin gene, reaction chambers or solid supports, reagents, solvents, probes, markers, detection tags, and the like.

It is also contemplated that each step of the disclosed methods can be in a separate kits. For example, there can be one kit for the amplification of amplicons of a nucleic acid sequence of interest and another kit for the sequencing of such amplicons.

J. Mixtures:

Disclosed are mixtures formed by performing or preparing to perform the disclosed method. For example, disclosed are mixtures comprising an amplicon from a nucleic acid

sequences of interest and a set of amplification primers. Also, disclosed are mixtures comprising an amplicon and a set of sequence primers.

Whenever the method involves mixing or bringing into contact compositions or components or reagents, performing the method creates a number of different mixtures. For example, if the method includes 3 mixing steps, after each one of these steps a unique mixture is formed if the steps are performed separately. In addition, a mixture is formed at the completion of all of the steps regardless of how the steps were performed. The present disclosure contemplates these mixtures, obtained by the performance of the disclosed methods as well as mixtures containing any disclosed reagent, composition, or component, for example, disclosed herein.

K. Systems:

Disclosed are systems useful for performing, or aiding in the performance of, the disclosed method. Systems generally comprise combinations of articles of manufacture such as structures, machines, devices, and the like, and compositions, compounds, materials, and the like. Such combinations that are disclosed or that are apparent from the disclosure are contemplated. For example, disclosed and contemplated are systems comprising automated delivery systems, such as robots, that deliver compositions, such as amplification primer sets, sequencing primer sets, reagents, solvents, and the like, to each of a plurality of reaction chambers or solid supports. Also, disclosed are reaction chambers or solid supports that contain or are associated with amplicons from a nucleic acid sequence of interest, i.e., a multi-exon gene. Also, disclosed are reaction chambers or solid supports that contain or are associated with amplification primer sets or sequence primer sets.

L. Data Structures and Computer Control

Disclosed are data structures used in, generated by, or generated from, the disclosed method. Data structures generally are any form of data, information, and/or objects collected, organized, stored, and/or embodied in a composition or medium. A nucleic acid library stored in electronic form, such as in RAM or on a storage disk, is a type of data structure.

The disclosed method, or any part thereof or preparation therefore, can be controlled, managed, or otherwise assisted by computer control. Such computer control can be accomplished by a computer controlled process or method, can use and/or generate data structures, and can use a computer program. Such computer control, computer controlled

processes, data structures, and computer programs are contemplated and should be understood to be disclosed herein.

The objects of the invention have been achieved by a series of experiments some of which are described by way of the following non-limiting examples.

Specific Embodiments

Disclosed is a method for characterizing a genomic DNA fragment by Single Condition Amplification/Internal Primer (SCAIP) sequencing comprising the steps of:

providing a PCR reaction plate wherein the wells of each plate contain the genomic DNA fragment;

adding to each of the wells a different set of left and right PCR primers complementary to a nucleotide sequence within the genomic DNA fragment and performing a PCR reaction at a uniform temperature;

purifying PCR fragments from each of the wells, adding the fragments to a corresponding well of a cycle sequencing reaction plate to which is added left and/or right internal sequencing primers corresponding to the PCR fragments, and sequencing at a uniform temperature;

purification of sequencing products followed by electrophoretic separation and fluorescent detection of nucleotides on a sequence analyzer; and

nucleotide sequence characterization.

Also disclosed is a method for identifying a mutation in a multi-exon gene by Single Condition Amplification/Internal Primer (SCAIP) sequencing comprising the steps of:

providing a sample of a patient's purified genomic DNA comprising the multi-exon gene,

plating the DNA in a 96 well plate followed by PCR amplification of gene-specific DNA fragments with a different PCR amplification primer set for each of the 96 wells under uniform amplification conditions, wherein each primer set is complementary to a single exonic region or a proximal promoter region of the gene,

cycle sequencing of the amplified DNA fragments with a different internal sequencing primer set for each well in a 96 well plate under uniform sequencing conditions,

electrophoretic separation of sequencing reaction products and fluorescent detection of nucleotides on a sequence analyzer; and

analyzing the nucleotides for mutations and comparing to other known nucleotide sequences.

Also disclosed is a method for diagnosing a distrophinopathy in a patient by Single Condition Amplification/Internal Primer (SCAIP) sequencing comprising the steps of:

providing a sample of the patient's purified genomic DNA comprising the dystrophin gene,

plating the DNA in a 96 well plate followed by PCR amplification of gene-specific DNA fragments with a different PCR amplification primer set for each of the 96 wells under uniform amplification conditions, wherein each primer set is complementary to a single exonic region or a proximal promoter region of the gene,

cycle sequencing of the amplified DNA fragments with a different internal sequencing primer set for each well in a 96 well plate under uniform sequencing conditions,

electrophoretic separation of sequencing reaction products and fluorescent detection of nucleotides on a sequence analyzer; and

analyzing the nucleotides for mutations and comparing to other known nucleotide sequences for the gene.

Also disclosed is a method for identifying a mutation in a multi-exon gene by Single Condition Amplification/Internal Primer (SCAIP) sequencing comprising the steps of:

providing a sample of a patient's purified genomic DNA comprising the multi-exon gene,

plating the DNA in a 96 well plate followed by PCR amplification of gene-specific DNA fragments with a different PCR amplification primer set for each of the 96 wells under uniform amplification conditions, wherein each primer set is complementary to a single exon or a proximal promoter region of the gene,

cycle sequencing of the amplified DNA fragments with a different internal sequencing primer set for each well in a 96 well plate under uniform sequencing conditions,

electrophoretic separation of sequencing reaction products and fluorescent detection of nucleotides on a sequence analyzer; and

analyzing the nucleotides for mutations and comparing to other known nucleotide sequences.

Also disclosed is a method for diagnosing a distrophinopathy in a patient by Single Condition Amplification/Internal Primer (SCAIP) sequencing comprising the steps of:

providing a sample of the patient's purified genomic DNA comprising the dystrophin gene,

plating the DNA in a 96 well plate followed by PCR amplification of gene-specific DNA fragments with a different PCR amplification primer set for each of the 96 wells under uniform amplification conditions, wherein each primer set is complementary to a single exon or a proximal promoter region of the gene,

cycle sequencing of the amplified DNA fragments with a different internal sequencing primer set for each well in a 96 well plate under uniform sequencing conditions,

electrophoretic separation of sequencing reaction products and fluorescent detection of nucleotides on a sequence analyzer; and

analyzing the nucleotides for mutations and comparing to other known nucleotide sequences for the gene.

The multi-exon gene can be dystrophin, SOD-1 NF-1, ATM, dysferlin, calpain, $\alpha\beta\gamma\delta\varepsilon$ sarcoglycans, collagen 6A1-3, Nebulin, and Titin. The PCR primers can be selected from the group of primer sets as shown in Table 1. The sequencing primers can be selected from the group of primer sets as shown in Table 2. The dystrophinopathy can be Duchenne Muscular Dystrophy (DMD) and Becker Muscular Dystrophy (BMD). The mutation can be a deletion, point mutation, frameshift, duplication or combinations thereof.

Also disclosed is a PCR primer set which recognizes a single exon or a proximal promoter for the dystrophin gene as shown in Table 1. Also disclosed is a sequencing primer set which recognizes a single exon or a proximal promoter for the dystrophin gene as shown in Table 2.

Also disclosed is a PCR primer set which recognizes a single exon or a proximal promoter for the CAPN3 and DYSF genes as shown in Table 6. Also disclosed is a sequencing primer set which recognizes a single exon or a proximal promoter for the CAPN3 and DYSF genes as shown in Table 7.

EXAMPLES

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how the compounds, compositions,

articles, devices, and/or methods described and claimed herein are made and evaluated, and are intended to be purely exemplary and are not intended to limit the scope of what the inventors regard as their invention. Efforts have been made to ensure accuracy with respect to numbers (*e.g.*, amounts, temperature, etc.) but some errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, temperature is in °C or is at ambient temperature, and pressure is at or near atmospheric. There are numerous variations and combinations of reaction conditions, *e.g.*, component concentrations, desired solvents, solvent mixtures, temperatures, pressures and other reaction ranges and conditions that can be used to optimize the product purity and yield obtained from the described process. Only reasonable and routine experimentation will be required to optimize such process conditions.

A. Example 1: Single Condition Amplification/Internal Primer (SCAIP) Sequencing Method.

The genomic organization of the dystrophin gene was assembled from contigs downloaded from the UCSC Human Genome Browser (Kent *et al.* (2002) Genome Res 12:996-1006) (see also the International Human Genome Sequencing Consortium 2001 (Lander et al. (2001) Nature 409:860-921)). Assembly and exon-intron annotation was performed using task-specific Perl scripts. The completed assembly reveals that the DMD region is currently contiguous and gap-free for the dystrophin Dp427m muscle isoform (NM-004006) spanning 2.09 Mb, and the dystrophin Dp427c brain isoform (NM-000109) spanning 2.22 Mb of chromosome Xp21.2. Primer systems for polymerase chain reaction (PCR) were designed to amplify DNA fragments which span each exon and 7 of the 8 promoters (Dp427m, Dp427p, Dp427c, Dp427l, Dp260, Dp140, Dp116) (Table 1). Each amplicon was designed for an optimal size range of 1.2 to 1.4 kb with the exon, including unique promoters, centered within the amplicon, with the exception of exon 79 which was broken into 7 fragments to maintain uniform conditions. These were designed to produce 93 amplicons with a nearly universal size; this uniformity allows one to predict likely amplification conditions using a single set of PCR temperatures.

Table 1. Primer Pairs Used to Amplify the DMD Exons and Promoters and Sizes of PCR Products.

Exon	Forward	Reverse	Product Length (bp)
M1	AATTGGCACCAGAGAAATGG	TCATGTGTTTAGTTCTATCGCAAA	1223
2	TCATTTCTCCATGGTTGGGT	TGACATCCCAATAAACCTCCA	1400
3	GCTCTCACAGGGTTGTTTCA	GAAGGGCAAAGATAAGAGACGA	1347
4	GGGAACCAAAGTGATTGAGG	TGCTTGGAGACAGCGTTTAAT	1367
5	CAGGAGACACAGAGATTTGCC	TCGGAAGACCCTATGCTCAG	1148
6	GCTTGCGTTAAATGATGGTATG	TTGATTTGCTGTTCCAGTGC	1391
7	GCGTAGATTATTTGTCATCTTCAGG	TGAGTAACCATCCAACAGAGGA	1245
8	TTATCCCATGCACCACAATG	CAAGCCAATGTCATGGAAGA	1360
9	CTGCTGAATGTGTGGAGAGC	ACATTTCATTCCCACGCTGT	1298
10	GGCCTTCTGGAAATAAAGGC	AAACTTGTGGCCCATTTAGA	1349
11	GCCAACAGGAATACGAAAGC	TTCAAATCCACAGTTGGCAC	1348
12	TGCAGAATCACTCCTATATGGTC	CATACCCTGCGTTGTTCTCA	1336
13	GGAGAACATCCTGCTGTACCTT	TAGCAAGGGCTTTCTCTCCA	1184
14	TTCCTTTGCATAGAAAGCATCA	AACCGTCGCTTGTAACTCTCA	1398
15	CCAAATGGTAGGCAATTCTCA	AATGTCAGGATAACCGTCGC	1148
16	CAGCATTTCAGAATGGCAAG	TGAAATCAGCAGTCTATGGCA	1177
17	TGTCTCCAGTGATGAATATGGG	TGCGCAGACTGAGACATCAT	1333
18	AAGCTCTGACATGCAAGCAC	ACTGAGAAAGGCTGGACACC	1171
19	TTGTCTTCCTTGGAAATAGGAG	TTTGGAAATAGCATTATCCCTGA	1399
20	ATTTAAACTAATTTCCAAGCCCA	ACACTATCCGGTGTGGTTCC	1232
21	GCCTGTTTGGTCAGGACAAG	GCTGAGTTTCAGTTGCCACA	1247
22	TTGCAATTGGGATTAACAATG	CCCACCAGTTTGAGAATGTG	1117
23	ATCCTTGAATCCCACCATAAT	CAGCAGAAATGAAAGGTAATATAGGA	1168
24	GGGAAAGAATCATGGGTGAG	CTTCCTGCTGCATGACAATG	1256
25	CATTGTCATGCAGCAGGAAG	ATGTGTCGAAGAGGCCAAAC	1081
26	TGAATTATCATCATCGGGCA	CCTTGTCACAATCCTTGAACC	1271
27	CACAAATCCATACCTCCATGC	TTGAGGCACCTGCTTTCTTT	1078
28	TCCATATTCACGATGATGTTTACC	GAGCTTGAATGATTAAATGTCAGAA	1338
29	GCGAGTAGGCACTCTCTGCT	TCTTGCACATTCTAGGAAATCAG	1380
30	GATCATGCAAAGCTGGTTGA	TGCTTTCCAACAATGCCATA	1347
31	AGTATCTGCCGGAAGCCAT	GCAAGTGCATCTTCACTTCATC	1398
32	CATGGTAGAGGTGGTTGAGGA	ATTCGGTGTTGTCTTGAGGC	1330
33	TTCATCCAAATTTATGGCTAGAAT	AGTTGAGCGAAGTGAGATGGA	1203

34	CTGAGAACAGGAGCACAGGA	GCTGTGTCATTTGGTGATGG	1324
35	GGGCAGTTTCTTATTTGTGGA	TACCACCATTGACAAAGGCA	1268
36	CCATACAGAAAGCCGTTTCA	GACAGGGCATCCTAACAGTCA	1242
37	ACTTCAACCTCTGTGACCCG	ACCCTAGACCGTGCAGAAGA	1244
38	TGCATCACCAACCAAACTGT	CAGAGGTGATGGCAGTGAAA	1399
39	GGTTTCAGAAATGAAGCAGGA	TCCTGCACAAACCAGATGAG	1290
40	AGCCTTGGAAGGAGAAGCAT	ATTCCTCTGGTGTCTTGGGA	1398
41	AGCCCATTCATTTCATCAGAG	ATGGCTTATGCAGGTTGACA	1155
42	GAAATTTAAATGCCGGTTGC	GCTTCCAGGAAACCATTTGA	1371
43	CACCATTTGCTACCTTTGGG	TTCAGCTCATTTGTCTGAATTG	455
44	GGATTAAAGAAGGCATCGCA	GGTTCCAACATAAAGCCGAA	1372
45	ATCTTGATGGGATGCTCCTG	CATTTGGCTTTCTGTGCCTT	1370
46	CAGATATAATGACATAATGTTGTTAGA	GCAATCCAGATCTTCCCTAAG	1264
47	GTCTTGGGAAAGGGCATACA	ATAGTATGCAAGGTGGAAAGATG	1374
48	CCTATAATCATTCTGTTACAGTCTAC	GAAGCCTGTCAGTTTACAAGAAC	1370
49	TGCTTTAAGTGTTTACCCTTTGG	CTGACCTGGCTTTCCATCTC	1247
50	GCTAGTTGCTGAGAGGGAACTG	AAGCCAGCATTAACATTGCC	1244
51	TTCATTGGCTTTGATTTCCC	GAAGGCAAATTGGCACAGAC	1198
52	GATGCTCTCCAAACTTGCCT	AAGTTCCTGCCCACCCTACT	1298
53	CAGAAACTAATATTTGCCATCAAAA	GAGAAGAATGAGCTGGGCTG	1162
54	AAGCCTCCTCTCTGCACTTG	CGAGTCATATTGCCCTCCAC	1378
55	AGCAGCATCAAAGACAAGCA	CGACAAATTCAGCCATCTCA	1159
56	GGCCAAGTGCAATCTTGTTT	TTCCTCCACGGAACTATTGC	1380
57	GGCTGCCTAGGGTGTAGAAA	TTGATTGCATGTTGAAATGAC	1375
58	TCCGCAATTCCTACATCCAT	GCTTTCGTAGAAGCCGAGTG	1399
59	ACCAGGAGCCCAGAGGTAAT	AGGGCAACACATTAACAGCC	1357
60	CCATTGTTATAATACTACCACAAGAG	GTGGCAATTCACATCTTCCA	1309
61	CCAAATTTAAGCCTTGCCTG	TGAACTGAACTGATAGGCAGAAA	1325
62	GCAAAGATCATTCATTTGACCA	AGTCGAGGACTGCTGCTTTC	1250
63	GCTTCATTCAGGCCCAAGTA	GACAAACCAGACATCTGGACA	1369
64	AGTTTATGGGCTTGTGGATGA	GCACACAGACCCTCAGACAA	1176
65	GCAGAGAGATGCTGAGGTGA	ATCTCCCTTGTGTGCAATCC	1350
66	TGTGTTATGTGGCCTGAAGTAA	CAACTGCAGCCTTTCACAAT	1275
67	CCTGTGGGAACACATACATGA	GCAATGGGACAGGATAGGAA	1192
68	CAGACAGAATCAACAGGGCA	TGGTGCAAAGTGAATGAGAGA	1242
69	TTTGAAGATGAATCGTATCAGTCAA	CTACATCCTTGCCATTTCCC	1335
70	TCCTCCCAGATATTTGCCTG	GGAAAGCAATAGCCAAACCA	1222
71	CTCTCAGCTGAACACCCTCC	CCTGATAAACAGTCCGCACA	1210

72	CTTGCTGCTGAATTGGAAGA	TTGACAAAGGATGGATGCAC	1318
73	ACTTGCCCTCTAACGTGCAT	GGCAGGTTTGGTCAAAGATT	1383
74	CTTGGTGGCCAAAGCATTAT	GGGCTCAACCAAAGAGATGA	1354
75	TGGCATTATCTCCTTGAGGG	TCCCAGAAAGCCAGAACTATG	1318
76	CATAGTTCTTTAAGCCTCCATCG	CCAACCAAATCCTCTCCCTT	1301
77	TGAGGAGACAGCACTGCAAG	AAAGGGCACCTCATAATTAACTCTT	1397
78	CTCTGTGGCTTGCCCATTAC	TGAAGGGTACGTTGAGATGATG	1389
79a	GAAAATAGCCACCTCCACCA	ATATCACGCCAAAAGGATGC	1153
79b	TCACTCATAGCCAAGGTGGA	AAGCAGGTAAGCCTGGATGA	1227
79c	TTACAACTCCTGATTCCCGC	TCACAAATGTGATGGGGCTA	1380
79d	ATATGGAACGCATTTTGGGT	CCTGTGTGGAACTACTCGCA	1206
79e	GCCAGGAGGAAACTACACCA	GGTCCAGCGTCACATAAAGG	1295
79f	ACTCCCAAGCAGTAGCAGGA	CATGCCATGTGATGTTTATGC	1319
79g	AGCCCATGAACTGTGTTTCC	AGCAATGAGGATGATTGATTGA	1376
Mp1	GGGCACTTATACTCTGGGCA	CGCCTTCTCTCTCAAGTTGG	1351
Mp2	TCAACTAAGGCTGAATGGCA	ATGCCCAGAATAATCCATGC	1370
Lp1	CCATATCTAGAAGCTTTATTCTGTTTT	GAATCTGCTTTACAGTGGTTGAG	708
Pp1	TGATCAGATGGGGATTGACA	TTCATTAAAGCCACAACCCA	1324
Cp1	GCATACAGGGTGCCAGACTT	TAGACCAGCTGGGTCGACAT	1399
260p1	CTCAGTCATGCTCTGTGGGA	ATCAAAACAACCCCATGGAA	1183
140p1	CAATAGCCCCATTGCTCAGT	AAGAGGGCACAAGCTTTGAA	1263
116p1	CGTTCTGCAAGAATCCCAAT	TCTGACCATAAAAGCGTGGA	1322

The primer sequences in Table 1 are SEQ ID NOs: 1-186, respectively (forward primer, reverse primer, from top to bottom).

Fifteen picomoles of each primer was aliquoted into individual wells of a 96-well tray, evaporated to dryness in a speed vac system, and stored in a -20E C freezer until use. For PCR amplification, 10 μg of patient template DNA was aliquoted into a master PCR mixture and subsequently 25 μl of the mixture was aliquoted into the 96 well dish with dry primers. The PCR was carried out in a thermocycler for 25 cycles under the following conditions: denaturation at 94° for 20 s, annealing at 55° for 30s, and extension for 68° for 4 min, followed by a final extension at 68° for 7 minutes.

To validate PCR amplification and to detect any deletions, 3 μl of the PCR product was run on a 0.75% agarose/Ethidium Bromide gel. The resulting gel was photographed and analyzed for absence of one or more bands. Because the absence of a single band may result from a primer site polymorphism, in such cases PCR was repeated using (1) the same primers, (2) internal sequencing primers, and (3) combinations of original and internal

primers. The absence of more than one adjacent exon is interpreted as being consistent with a multiexon deletion. The PCR products were then transferred and bound to a 96-well filter plate (Millipore MAFB 1.0 :M glass fiber type B filter) in the presence of a 5 M guanidine HCl/potassium acetate solution. Wells were washed four times with 80% ethanol to remove unincorporated primers, nucleotides, and excess salt, followed by elution of the fragments with warm nanopure H_2O.

Internal sequencing primers were designed to anneal to unique intronic flanking sequences, with attention to specific 3' sequence for each primer (Table 2). As with the PCR reaction, the primers were stored in 384 well plates so that both PCR set-up and sequence reaction set-up could be performed with multi-channel pipettors and pipetting robots.

Table 2. Internal Primers Used to Sequence the DMD Exons and Promoters.

Exon	Internal Primer A	Internal Primer B	Primer Distance (bp)
M1	CACTGTGCTATTCTGGTTTGGA	TTTATGCTTCTTTGCAAACTACTG	595
2	ATTTTAATTTGGATGCCCCA	TCTTCTTCTGCTGGGTGACA	563
3	TTACTCTTGCTATTCAAACTAATTCAA	TTTTCTGCAGGCGGTAGAGT	501
4	GCTAAAAACGTACCAGGCCA	GGAGCAGCCTATCAGGTCAG	503
5	TCCAGTTGACCTCTTTAATCTGC	CCGTGATGATCCTTAACATTTC	516
6	TGGCATAGATACCAATGAATCAG	TGTATCCCATAGAACACTGGAAAA	562
7	AGGACTATGGGCATTGGTTG	TTTTCCTAAAAGTCTTCACTGCAA	461
8	TGCTCATCTCATTGGTCTGC	CAATGAAGCAAAATTGAAAAGG	560
9	AAGTGCCTTCATTCTGGGAG	GAAACCATTACGGGAATTCAT	542
10	GGATTTTGACCGCTATTTGAA	GTTGGCCGATCAGGTAGAAA	595
11	GTGGTTTTGGGATTCTGCAA	CAGTGCATCTATCTAACATCTGCTC	548
12	AATAGTTCCGGGGTGACTGA	GGAGGGGACTTATTCAAGCC	509
13	TGGCTTGGAATGGTTTTAGG	GATTTTACCCATCCGCAGTT	475
14	TTGCTTGTCTCTTTGCTTTTC	CATACGGCCAGTTTTTGAAGA	547
15	TCGATGGGCAAACATCTGTA	TTGAAAAACAAAGTTGAAAATCCA	505
16	GAACTTTTGATCCTTTGCGG	TCACCACCATTCTCCAACAA	493
17	TGTTGAGATTACTTTCCCTTGC	TTGCGATAGTGATTTCTTGTGA	571
18	AACAGGGAAAATAGTGCTGCT	GGCATCCCTAGTCAGTCACAG	491
19	TCATGAAAATGGCTCATGCT	CCACATCCCATTTTCTTCCA	497
20	TTGTTGTGACGCAAGTCTGA	TTGCGCTTAGCTAAATCCTT	565
21	GGCTGGTGATAGAGGCTTGT	TCACAAAATTATTATGAGGACAAAAA	544
22	ATGTGTAAGGTCCCTGGCAT	TTTTCATTTGCTCAATGGG	475
23	TCAGAAAAATACATATGGAGTGTTAAA	AAGGAATAAGCAAATCGCCA	612
24	GCCTCAAGAACTACTTAGAGACATCC	AGGCAATGTTTTGTCAGTTCC	581
25	CCCACTGGATTCATGCCATA	TTTTAGGATCAAAATAAGATGAATGTG	583
26	TGAGTGTATCTGATCCCCATGA	TCTGATCCCCATGAGTTATTTTC	

27	TTTATGGAAGAGACTGGAGTTCA	GGAGAAAATTTATAGGATTTTATGACC	672
28	TTTCTTAATGACTTTTGATTGTAGAGG	GAAGCCATTTAAACCCTTTGC	534
29	GCAAAAATGCTCCTTGGTGT	CAGTGTCTGGCATTGGATTG	446
30	GGAGGAACATTCGACCTGAG	TCCTACCTACCTCCAAATAGTCAAA	638
31	CCCATAGGGAAGAAATAAATCG	CATACATTTGGGAGAATGATTCAG	618
32	TCCTGTGTTGGATGAATGGA	GCCACAATACATGTGCCAAT	483
33	ACCGCTGCAAAATGCTACTC	CTGAATAAGCAGAGCCTCACTG	557
34	ACGATGTCATCTGCCCTAGC	TCATGGTCCTGAAAAGCACA	526
35	TCATAGTTACCCAACAATGAAGC	AGTTTCATTGAGATTAGTTTTAAGTGG	574
36	CGCAATATTCTATATGAAAATACCACT	TGAGTGATGGATTTGAACAGAAA	487
37	CCCTTTGTATTTTCTGCATGTG	GGGAGGAGTGGCGTTTATCT	517
38	TGCATGTATGTTCAGCTCTGG	TCAAAAGAAAATTGCTGGGC	578
39	CAGGTGCCCCTAAAAATGTG	GCAACACATCGTTCAAAATCA	553
40	CTTCCTATACATGGGTCCCG	CAAGGAAATGCATCAAATCAAA	471
41	GGGTTATTGAGCGAGGATGA	AAGCCCAAAGTGAGGGAAAC	506
42	GCTTTTAACACTTTCTGGAAAAGTAAG	AGATTTCTGAAGCCAACCACA	558
43	CACCATTTGCTACCTTTGGG	TTCAGCTCATTTGTCTGAATTG	455
44	TTGTGTGTACATGCTAGGTGTG	CCAGGCAAACTCTCTCATCC	541
45	GGGAAATTTTCACATGGAGC	CCTTTAAGCAATCATGGGTGA	571
46	TGAATCAGAATTTTTCTTGTTCGAT	TAAGCGCTAGGGTTACAGGC	-
47	GAGGGGGTGAGTGTTTCAGT	AAAGCCATTCACCATCATCA	532
48	TCAGTTGCAGTTGGCTATGC	GTGAGGTTGGTTTAGCC	809
49	TCTGTTTCTTTTCTCTGCACCA	GAGTCCTTTAAAGCAATGACTCG	487
50	TATTTGATGGGTGGTTGGCT	CCGTTGTCATGCAACACTTT	490
51	TCATGAATAAGAGTTTGGCTCA	TTAGGCTGAATAGTGAGAGTAATGTG	522
52	CGGAATGTCTCCATTTGAGC	TGCTTTGCAACTATATAAGCCC	605
53	TGTTGTTCATCATCCTAGCCA	AGCCTGGGTGACAGTGAGAC	507
54	TTTGTCCTGAAAGGTGGGTT	AGAAGTCTGAGCCAAGTCCG	506
55	TGTCATTCTTGCATGCCTTC	CCTCCTTGTCCAAATACCGA	565
56	CAATACGCCAAGAAAAGGGA	TGATGTCTTAATATGCATGTCTCC	589
57	CCTCTGTTTTGTGGCTCTCA	GCCAAAAGAGATGGACGATT	531
58	AACACAGCGCTTTCCTCATT	TTCCTCCTCACAGATAACTCCC	595
59	GGGCTGTATCAAAATTTATGCC	TTGTGGGAAGATAACACTGCAC	514
60	ACTGGCACTGCACCCTAAAG	AATTTGAAAATGTTTAGATGGGAA	410
61	ATCCTTTGTGTTTGGCCTTG	ATCCAATTGGCCTTCCTCTT	475
62	CGCATTTATCTTTGTGCCTG	CGCAAAGATTGACTCCCACT	587
63	GGGCCTTTCTGCTTGTAAGA	CAAAGACCTATAGGCCCTCTCA	489
64	GTTGTCAAAGGGCAAAAGGA	AGCTGAGGAATGGTGACAGG	492
65	TGTGGTTCACGTTTGGTGTT	GAGAGCAATCTACATTCTGGCTC	529
66	TGGTTGAATTTCCATTGCAT	TTGACAAGGAATGGCACAAA	470
67	GCACAAATTAGAAGTAACCCCA	CCTGCTGCAGATGGAGATTT	520
68	AGCTGTGAAAAGCCAGCCTA	GGGTAGCTCTTTGGATATCAGG	-
69	AGCTGAGTTTTTCTTCCCTCC	GAAGCCTACAGTTGAGAGCCA	500
70	TTGAGTAGCCTAGTAAGCTTGTATGT	AAAGTGGCAACTGGACATCAG	596
71	GATCAAAGGGGACGTCTTCA	ATGTCCAGTTGCCACTTTCC	-
72	CGATGGGAATTTTCCAGAGA	CCGGAAATGTTTAAAAGCCA	554
73	TGGTCTACCACACACTGCCT	AAGATCACGTTTCCACTCCC	643
74	TGGTAGATCACAACCTCAGCA	CTGCAAATGGAGCTAAACAGA	469
75	GCCTCTTTTGCTTGCTGTTC	TCACTTTGCAGGCACATACC	522
76	GGGAGCACAATTCAGATACAAA	ACAAGTTTTCTGTGGGCCAG	527

77	TGTATGGATTTCTTCTTCCCTTT	GAAACATGTTGCCCTCACG	482
78	GCTGCAAGTGGAGAGGTGAC	GGGACTACAAAGGATTGCCA	-
79a	TTCTTCCTGGAAACTGGTGAA	GCACACTTTAGTTTACAATCTTTCTTT	599
79b	AACAATGGCAGGTTTTACACG	AAGCAGGTAAGCCTGGATGA	581
79c	GGCAGGCTTGAGTTTTCATT	TACTCCTTCACAGGGATGGG	584
79d	ACATTCAGCTTCCTGCTGCT	AACCTGTCTAATCCACCAAGAA	573
79e	CAGGTATCAACCCAGAAGCC	GAGCTTTGGGTTTTCTTTTGAA	600
79f	TTTGGAGAGTGGGCTGACAT	GGTGGTTATAAAGAACACAACACG	599
79g	AAATCAGAGGTAAATAGAGTGCATAAA	GGGGAAGGGGTAGTTAGGAG	597
Mp1	TACTCATTGCAGTCGCAAGC	TGATGATGCCAACAGTCTGAA	581
Mp2	GCATAATTCACAACTGAAATTTAGGA	GTAGAGGCCCCCGGATATT	654
Lp1	AAAACAGAATAAAGCTTCTAGATATGG	GAATCTGCTTTACAGTGGTTGAG	708
Pp1	GGTGTCTTCATAATAATCAGCTCC	CTCACAACAAAAGCCCCAA	658
Cp1	TCAGCCAAAATTTCAGTGTG	GCAGAGTTTGAAGAGCTCGG	637
260p1	CCAATAAGTTGCCTGCCCTA	TGTGAAGGAGAAAAATAAATAGCAAA	637
140p1	TCAGCAAACCTTGCATTTTT	CACGCTCCTGCATCAGAATA	674
116p1	CAAAGCCTCCATTCATTGT	TGATTTCCCATTTAATACACATTTTT	610

The primer sequences in Table 2 are SEQ ID NOs: 187-372, respectively (internal primer A, internal primer B, from top to bottom).

The sequence reactions were assembled by transfer of a uniform concentration of PCR product to a new cycle sequencing plate along with 10 picomoles of sequencing primers, and the samples with primers were evaporated to dryness in a speed vacuum system. The fragments were rehydrated with a mixture of ABI PRISM BigDye terminators v.3.0, the plates heat-sealed with a foil seal, and placed on thermocycling blocks for cycle sequencing. Post-cycling processing involved ethanol precipitation in the cycling plates, rehydration in formamide and re-sealing. The plate was then placed on the plate deck within the ABI 3700 for robotic loading, capillary electrophoresis, and fluorescent detection of the sequence ladders. All plates within the system were bar code labeled with plain sample identifiers. These bar codes were captured at multiple steps of the process using a web-based system for plate tracking.

1. Sequence Analysis.

After initial data processing using ABI 3700 instruments, sequence trace files were transferred onto a Linux disk server. The base calls were reanalyzed with the Phred program (Ewing *et al.* (1998) Genome Res 8:175-185) that adds a quantitative base quality value. This base quality value provides a probabilistic estimate of the correctness of the base call. The quality values are the log of the probability that the base call is correct, such that a Phred value of 20 corresponds to a 99 % probability that the base call is accurate, while a Phred value of 30 corresponds to a 99.9 % probability that the base call is accurate.

The sequence was assembled with dystrophin consensus sequence using the Phrap program, and potential mutations were identified using the Consed program. The read assembly was performed on a PCR fragment basis, and a single PCR Phrap assembly consisted of the consensus genomic sequence and all sequence reads relating to the PCR. The read sequence and Phred quality values were compared to the assembled consensus sequence using cross_match, and all discrepancies were tagged and ranked depending on Phred quality of the base (cutoff of 15). All PCR assemblies (Reads + consensus sequence and tagged discrepancies) were then compiled into one consed project for review. Potential base discrepancies were catalogued using Perl scripts, and underwent human review of original trace files. This final list of reviewed discrepancies was loaded into an Oracle database where they were further reviewed in a web browser.

Nucleotide sequence position was based on the annotated mRNA sequence found in GenBank (NM-004006) which encodes the dystrophin Dp427m isoform.

B. Example 2: Description of DMD Patient Population Used in SCAIP Sequencing Analysis.

Patients from the University of Utah's Muscular Dystrophy Association clinic were ascertained for disease status. The diagnosis of a dystrophinopathy was determined by the presence of clinical features consistent with Duchenne (DMD) or Becker (BMD) muscular dystrophy, along with either (1) absent or altered dystrophin expression by immunohistochemical or immunofluorescent analysis, or immunoblot analysis; or (2) a clear X-linked family history. Some patients had previously had confirmation of dystrophin deletions by clinical testing. Probands from 42 families were enrolled. Forty-two were males with dystrophinopathy by the above criteria; the forty-third was an obligate carrier female (and the mother of two deceased Duchenne patients) with adult onset limb-girdle weakness which led to wheelchair dependence in her sixth decade. Nine additional DNA samples were obtained from self- or physician-referred patients nationwide who had been shown to be deletion-negative on standard screening.

Patients were catalogued as to whether they harbored large-scale dystrophin deletions detectable by standard clinical multiplex PCR analysis. Blood samples for DNA analysis were obtained under an IRB-approved protocol from patients who either had no clinical record of dystrophin deletion testing (unknown deletion status) or who had no detectable deletion by commercial testing. DNA was obtained from each blood sample using a salting-out method (PureGene, Gentra Systems, Inc; Minneapolis).

Direct sequence analysis was also performed on 66 DNA samples from one clinical center (O.S.U.). Sixty-four of the samples had previously been evaluated by the DOVAM-S technique. Clinical phenotype of this set of patients was confirmed by clinical exam and muscle biopsy.

SCAIP detected dystrophin mutations in 70% of patient samples which did not have deletions of more than one exon. Excluding five patients with duplications from the Utah/referral set, the detection increased to 74% (62/84). This is probably an underestimate of the actual rate of detection in the general non-duplication sample population, as duplication testing was not performed on the DOVAM-negative/ SCAIP-negative set (n=17).

Correlating these numbers to the general dystrophinopathy population is unhelpful, because the patient set was not a random sample; it likely represented a population enriched in duplications as well as stop codons and subexonic rearrangements. The absence of detectable mutations in the remaining patients is not yet explained, but unlike the case when DOVAM or DHPLC screening is performed, the known coding regions of the dystrophin gene do not contain disease-causing subexonic mutations.

C. Example 3: Large scale (≥1 exon) deletions.

Deletion status was determined by reviewing clinic records or obtaining clinical (multiplex PCR) testing in 42 Utah probands. Of all the samples, such deletions were found in 25/42 (59.5%) patient samples. As discussed below, a single Utah sample had a non-hotspot single-exon deletion, bringing the total found in the Utah cohort to 26/42 probands, or 62%.

D. Example 4: Direct Sequence Analysis by SCAIP Sequencing.

1. Amplification efficiency and deletion detection

In anticipation of direct sequence analysis, PCR amplification was performed on 94 samples. These included the remaining 17 Utah probands without multiplex deletions, and 9 referral samples (total unique families n=26); two relatives of Utah probands (1 asymptomatic carrier mother, and 1 affected sibling); and 66 samples from O.S.U. (64 DOVAM-screened and 2 unscreened). PCR amplification was performed on a total of 94 specimens. An aliquot of each well from the 96 well PCR amplification plate was loaded in 96 well format onto an agarose gel. Electrophoretic separation distance for each band was ~1.8 cm, as the wells were angled slightly relative to the migration path. The products were from a multiexon deletion case missing exons 20 to 30 and the DMD260 promoter.

Products corresponding to exons 1 to 78 are located in sequential wells, starting left to right and top to bottom, followed by the multiple exon 79 and alternate promoter products. Note the absence of products in wells corresponding to exons 20 to 30 and Dp260.

Analysis of PCR products by visualization on agarose gels resulted in the identification of three individuals with deletions of ≥1 exon as shown in Figure 1. In one OSU case, multiple amplification products from adjacent exons (the DMD260 promoter, and exons 20-30) were missing; review of records (unblinded only after the entire sample set was analyzed) showed that this had been detected by DOVAM analysis. In two patients, single amplification products were not present in exons not screened in commonly-used multiplex screening sets; in each case, PCR was repeated with internal primers in order to exclude the presence of polymorphisms at the primer sites, and the absence of a product on the second round of amplification was interpreted as representing single exon deletions. One Utah patient had a deletion of exon 18. One OSU patient had a deletion of exon 21; unblinded post-amplification review of the DOVAM results showed that a possible deletion had been suspected, but that a primer site polymorphism could not be excluded. The overall efficiency of PCR is summarized in Table 3.

Table 3. Efficiency of PCR Recovery.

PCR recovery	94 individuals x 93 PCRs = 8742 PCR potential products	
		efficiency
Primary amplification:	8716 / 8728 Total exons = 8742 – 14 deleted exons = 8728 potential products	99.86%
Primary sequencing:	8396 / 8449 Three deleted samples not sequenced = 93 X 3 = 279 exons Total exons = 8728 – 279 = 8449	99.37%

Excluding exons determined to be deleted in these three patients, the efficiency of primary PCR recovery (defined as the presence of a band on first pass, single plate amplification) was 99.86%.

2. Sequencing efficiency and quality.

Direct sequence analysis was performed on 91 individual samples. The overall quality of sequence recovery is shown in Figure 2. Each block represents the length of the individual PCR products, with the exonic sequence indicated by the thick line on the top horizontal axis. The average Phrap score observed in this study is plotted along its horizontal position, with the vertical axis ranging from Phrap score 15 to 50. Phrap scores >

50 are not shown, and the portions of the plot corresponding to the exons +/- 100 nucleotides are indicated in gray. The Phrap score over coding regions of the gene is generally > 60. The efficiency of primary sequencing recovery (defined as high quality sequence on the first sequencing reaction) was 99.37%.

E. Example 5: Mutation and Polymorphism Detection.

Among the samples from the 16 Utah probands and 9 referral samples, mutations were detected by SCAIP sequence analysis in 16; five additional samples harbored duplications (see below), resulting in an overall detection efficiency of 80% in this group (16/20 non-duplicated patients). The mutations are summarized in Table 4. These include ten stop codon mutations; one single base pair (bp) insertion; and one single bp deletion. The single base pair insertions and deletions were easily detectable as mixed base calls in the two females tested.

In two referral samples, sequence variations were detected that may be causative of disease by altering intronic splice signals. One sequence variation is highly likely to cause disease, as it occurs in the highly conserved +1 position in intron 25 (changing a G to a C). The other is less definitively causative, as it occurs in the less conserved –9 position in intron 11. Both are unique in our series (n=94) and are previously unreported, according to the Leiden database of dystrophin mutations (http://www.dmd.nl/dmd_all.html). Definitive assignment of a causative status to these two will sequence variations will require analysis of dystrophin transcripts; muscle samples are at present unavailable, although further studies are planned.

Of particular interest are two substitutions which result in nonsynonymous changes in amino acid sequence in highly conserved functional domains of the dystrophin protein. One of these, in a boy with a DMD phenotype (loss of ambulation at age 10 years) substitutes a phenylalanine for a cysteine in the dystroglycan binding domain, in a residue conserved in the dystrophin protein through *C. elegans*. The second, in a boy with a BMD phenotype (still ambulant at age 16 years) substitutes a valine for an asparagine at a similarly conserved residue in the actin-binding domain.

After direct sequence analysis was performed, dystrophin duplication analysis was performed in 13 samples, including the 9/25 Utah or referral samples without detectable mutations, and the four with presumed mutations discussed above (two intronic and two missense). Duplication analysis was performed using the multiplex amplifiable probe hybridization (MAPH) technique (White et al. (2002) Am J Hum Genet 71:365-74). No

Table 4

	ID	No.	Type	Age at: Presentation	Age at: Loss of ambulation (current age)	Mutation: Exon	Mutation type	nucleotide	amino acid	novel
Utah Non-deletion, Non-duplication Samples (n=12 probands). ***Mutations (n=9 probands)***										
Stop codons	42172	1	DMD	15 m.	9 y	47	Stop	6868A>T	Lys2290X	+
	42588	2	BMD	3 y	n.a. (10 y)	31	Stop	4250T>A	Leu1417X	-
	42719	3	BMD	13 y	n.a. (19 y)	31	Stop	4240C>T	Gln1414X	+
	42953	4	DMD	6 y	9 y	64	Stop	9337C>T	Arg3113X	-
	42970	5	BMD	20 y	n.a. (58 y)	1	Stop	9G>A	Trp3X	+
Deletions	42390	6	DMD	3 y	n.a. (4y)	30	1 bp deletion	4103delC	frameshift	+
	42389	*6a*	*mother of indiv. 6*	*asympt. carrier*	n.a.	30	*1 bp deletion*	*4103delC*	*frameshift*	+
Insertions	42359	7	Manifesting carrier (female)	30 y	n.a. (58 y)	8	1 bp insertion	783_784insT	frameshift	+
Missense	42458	8	DMD	5 y	11 y	68	missense	9938G>T	Cys3313Phe	+
	42515	9	BMD	6 y	n.a. (16 y)	6	missense	494A>T	Asp165Val	+
No mutations (n=3 probands)	40818	10	DMD	7y	10 y		n.d.		None	
	42273	11	BMD	8 y	n.a. (18 y)		n.d.		None	
	42965	12	BMD	13 y	n.a. (21y)		n.d.		none	
Referral Samples (n=8 probands)										
Mutations (n=7 probands)	42962	13	DMD	4 y	n.a. (5 y)	53	Stop	7720C>T	Gln2574X	+
	42964	14	DMD	4 y	n.a. (7 y)	34	Stop	4693C>T	Gln1565X	+
	42968	15	IMD	2.5 y	n.a. (13 y)	58	Stop	8608C>T	Arg2870X	+
	42969	16	BMD	3 y	n.a. (11 y)	5	Stop	355C>T	Gln119X	+
	42971	17	BMD	5 y	n.a (21 y)		splice		IVS25+1G>C	+
	42974	18	DMD	4 y	12 y		splice		IVS11-9G>A	+
	42986	19	DMD	2.5 y	10 y	34	Stop	4693C>T	Gln1565X	+
No mutation (n=1 proband)	42963	20	BMD	5 y	n.a (11y)		n.d.			

duplications were detected in the samples with the four presumed mutations. Of the remaining nine samples, duplications were found in five (data not shown). Of the four remaining patients without detected mutations, one patient (#42965) was reported to have dystrophin of an increased molecular weight on commercially-obtained immunoblot analysis, raising the possibility that a duplication remains undetected by the MAPH technique.

F. Example 6: Comparison of Assay Sensitivity between SCAIP and DOVAM.

The SCAIP method was used to study 66 samples from a second center in a blinded fashion. Sixty-four of the samples had previously been studied by DOVAM, which identified subexonic mutations in 44 of the samples, and possible exonic deletions in two (discussed above). SCAIP analysis detected all 44 mutations as well as a previously undetected stop codon mutation (Glu2035X in exon 42, GAG::2035::TAG) in 1 of the 20 other non-deleted samples. This position is 2 nucleotides 5' of a common variant GAT::2035::GAG (Asp::Glu) that may have interfered with the SSCP analysis used in the DOVAM test.

Table 5. Summary of mutation detection in non-deleted, non-duplicated probands.

	# mutations detected	# samples	
Utah samples/referrals	16	20	80%
DOVAM positive samples	44	44	100%
DOVAM negative samples	1	18	5%
DOVAM unscreened samples	0	2	0%
Total:	62	84	74%

G. Example 7: Phenotype/Genotype Correlations.

The rapid and economical detection of stop codons and small rearrangements will facilitate the study of sequence context effects on disease expression. However, in the present study, only limited correlations between phenotype and genotype are to be drawn, although the results raise several interesting examples. One patient with BMD, the mildest affected patient in the Utah group, who is still walking at age 58 years, has a mutation

resulting in a premature stop signal in the third amino acid of the muscle isoform; the next methionine is at position 124. Another intriguing result is the presence in the relatively small sample size of two stop codon mutations in exon 31, both resulting in the BMD phenotype. Although stop codon mutations are expected to be essentially randomly distributed across the gene (unlike the hotspots found for exonic deletions) (Roberts *et al.* (1994) Hum Mutat 4:1-11.), the presence of two exon 31 stop codon mutations raises the possibility that stop codons in certain exons may predispose to a milder phenotype, perhaps due to the influence of such mutations in promoting exon skipping as seen in the *mdx* mouse (Wilton *et al.* (1997) Muscle Nerve 20:728-734; Lu et al. (2000) J Cell Biol 148:985-996). The mRNA and protein sequences in these and other patients have yet to be determined.

Two patients had a previously undescribed Gln1565X mutation. These patients are not known to be related, and analysis of single nucleotide polymorphisms (SNPs) reveals different haplotypes over at least a portion of the dystrophin gene, supporting the idea that they are unrelated, although distant relatedness with intragenic recombination cannot be excluded. This example illustrates one of the additional advantages of SCAIP analysis. That is, SNPs are found throughout the gene; some are quite common, others less so. Compared to screening strategies such as SSCP or DHPLC, SCAIP analysis allows one to detect a sequence variation with a greater degree of certainty, and the frequency of such variations can be readily established by comparison to the large and growing database of specific polymorphisms. By cataloging the SNPs throughout the coding and control regions for the dystrophin gene and establishing a rigorous and standardized phenotyping process, one is now enabled to generate testable hypotheses regarding the role of such SNPs on the presentation or progression of disease. For example, polymorphisms in the primary cardiac or brain isoform promoters could conceivably alter the clinical expression of cardiomyopathy or cognitive dysfunction. Studies to address these possibilities are underway.

H. Example 8: Implications for Clinical Use Including Genetic Counseling.

Application of the SCAIP method to the study and clinical care of dystrophin-related diseases will obviate the need for muscle biopsy in a large number of patients. It will routinely allow rapid detection in an economical fashion of the following gene variations in dystrophinopathy patients: (1) all deletions of > 1 exon; (2) small rearrangements of <1 exon in size (deletions and insertions); (3) premature stop codon mutations; (4) splice signal site mutations; and (5) missense mutations. Reports of non-synonymous polymorphisms as

disease-causing missense mutations in the dystrophinopathies are rare. Analysis of data generated by the present method will allow identification of variants at highly conserved amino acids in patients without any other sequence variation, leading to identification of greater numbers of missense mutations.

The availability of rapid direct sequence analysis will have an immediate impact upon genetic counseling in the dystrophinopathies. Because approximately one-third of all dystrophinopathy patients harbor de novo mutations, X-linked family histories are often absent, and testing of both known and presumptive carriers can, at present, only be performed with high reliability if a proband's specific mutation is known. In the absence of large-scale deletions, carrier testing relies on haplotype analysis. The high quality sequence acquisition method described herein allows ready identification of point mutations or small-scale rearrangements in the heterozygous state, and will lead to improved genetic counseling for dystrophinopathies as well as for other diseases to which it is applied.

I. Example 9: LMGD2A and LMGD2B Detection.

Limb-girdle muscular dystrophy type 2A (LGMD2A) is an autosomal recessive disorder caused by mutations in the CAPN3 gene, which encodes the skeletal muscle-specific calpain (calcium-activated neutral protease) (Richard et al., Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. Cell. 1995;81:27-40). Mutations are found throughout the CAPN3 gene and include nonsense, splice-site, deletions/insertions, and missense mutations (Richard et al., Calpainopathy-a survey of mutations and polymorphisms. Am J Hum Genet. 1999;64:1524-1540). There is some evidence for founder effects, however most mutations observed are "private" within affected families. LGMD2B is caused by mutations in DYSF, encoding dysferlin, a skeletal muscle protein associated with the sarcolemma (Bashir et al., A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. Nat Genet. 1998;20:37-42). PCR and sequencing primer systems for SCAIP analysis were developed for both the CAPN3 and DYSF genes. The PCR primers are shown in Table 6 and the sequencing primers in Table 7.

Table 6. Primer Pairs Used to Amplify the CAPN3 and DYSF Exons and Promoters.

GENE_EXON	FORWARD	REVERSE
CAPN3_1	GCAGTTCTCAGCTTCTTTCCA	GCTCTGTCATGTGCCCACTA
CAPN3_2	CTGCCCTAACTCTCAAGTTGC	ATTGGTTTGAAGGTCCCAGA
CAPN3_3	TTCCAAGGAAAGACTGGCTG	ACCAGCTCTATGCCAAGGTG
CAPN3_4	TCAATGAGGGAGAAAGTGCC	GTTGAGGAAGGGCTGCATTA

CAPN3_5	GCATTGCAAGTCTTGGATCA	TCAATATACTGAGCAGCCCTC
CAPN3_6	AGCTCCAAGTGTCAGGAAGC	TCAGTATTCTCCAGTGAGCAGG
CAPN3_7	CTCCTTAGGCACGGTCATGT	CACGAGAGAACAGGAAGCTCA
CAPN3_8	GCTTCCTGTTCTCTCGTGTTC	CTTCCACTCCTGGCCCTT
CAPN3_9	CCTGGTCTCAGGAATCTCCA	GAGAGAGGGTGAGGTTGACG
CAPN3_10	TCAGAAGTGACAGCGTTTGC	TCCTTCCCTACATCACCCAA
CAPN3_11	TGGCACTTGGTGATATGATAAGA	GTGCGAGGGAGAAAGTGC
CAPN3_12	AGAGAAATGCCTGAATCGTG	AGAAGACCCGGAGGATGAAT
CAPN3_13	TTGTGGGCAGGACTGTGATA	GTGTCACCAGAAGCAAGCAG
CAPN3_14	CTGAGCCACTGGCCACATTA	GACTTTGGGCTTCTCACTGC
CAPN3_15	AGGTCAGTTTGAGAGAGCCAT	TGTGGGTCTGGACAACACAG
CAPN3_16	TATCCTTGTCACTTGCACGA	AAGCTGGTTCTGTCTCAGCC
CAPN3_17	GGCCTTGAGCATTTCACAAT	CTCCTTAAGTTTCCCTGGGC
CAPN3_18	GGCTGGAGAGGTGTGAAGAG	GCTTTCCAGAGCCATCTGTC
CAPN3_19	GGCAGCTCTGATCAGGAAAG	TTGACTGCATTTCGCATCTC
CAPN3_20	TGAACCATGACCCTCCTCTC	GATGTGCAGGCAGAGAATCA
CAPN3_21	GACCTGAAGACACACGGGTT	CGCACTCCGCCTCTACTACT
CAPN3_22	CCTGGGTTACAGAGTAGGCG	GCAGCCACTGAAAGAAGTCC
CAPN3_23	GAGATGCGAAATGCAGTCAA	TCTGCAGACAGCCTAGAGCA
CAPN3_24	ATGGCAAAGGGAGGGTTACT	CCCGTTGTACATGACCCATT
CAPN3_EP1	CAGCGAACACTGGATTCTGA	TGGCTCTCTCAAACTGACCTAA
CAPN3_DP1	TTGTGGGCAGGACTGTGATA	GTGTCACCAGAAGCAAGCAG
DYSF_1	GCTGCCAAATACCCAAATGT	TCTGAGAGAGAGCAAAGGGC
DYSF_2	TTCTGGAGATGGATGTTGTTC	TCCCAACTCAGTTTCAACCC
DYSF_3	GGTGCTCAGGGACTCTCTTG	GCAGGTTGGGTTGAACTTGT
DYSF_4	TGTCAGTCAGAAATGCAGCC	AGGGCGGAAGTAGTTCCAAT
DYSF_5	TGTCACCAGTCCCTCTCCTC	CTGAGACAGGCACAGCACTT
DYSF_6	ATGGAGGTGCAGTAGGTTGG	GCTTGAACAAATTCAAATTCCA
DYSF_7	TCATCCATCTTCCCATTGCT	GCGTGTGCACTGACACCTAT
DYSF_8	GAAGCCAGTGGTGAGATGGT	CATTCACAGGGAACATGTGG
DYSF_9	TAAACTGCTAGGCGTGGAGG	TGGATCATTGCCTGTGATGT
DYSF_10	TTCTGAGAACCCAAGGGTTAAG	CAGCAGCCACTTCCTGAGAT
DYSF_11	TACAGAGAGCCCCGTGAGTT	AGCCATCAGCCATATTCAGG
DYSF_12	CATCAATGCATGTGGGATGT	GTCTAGTATCGGGCCAACCA
DYSF_13	TGTGTTGAATTCCCTGCAAC	GGTTCGGAGAGCTACGGAGT
DYSF_14	TTGGATCTGGTTTCCACTCC	CTTTCTAAGACGCCCGTGAG
DYSF_15	GAAAGCTGGTCTGGACTGGA	CAACTAGCAGGAGGTGGCAT
DYSF_16	TCTGCATAGGATGTGGTTGG	GAAAGGTCTCGGAGTGCTAA

DYSF_17	TTGTGGACAGTGTCTGGCTC	AGGTCATGCACTGTGAGTCG
DYSF_18	TTAGGGCAGAGGGTATGTGC	ATGACACCTCAAGGCCAGTC
DYSF_19	TGGATGACTACCTGGGCTTC	GGCAGGAACTCAATCCTACG
DYSF_20	CGTAGGATTGAGTTCCTGCC	AGTAGTGGCACCCTGGAATG
DYSF_21	CTGTTTGCGGCCTTCTACTC	TCTCCTTGCACTGGACACAG
DYSF_22	GACAGTCCTTGGCCTCTCAG	TTAACCCTGTGGAGAGCAGA
DYSF_23	TTCTGGGAAGGGTTCTGTTG	GAGCAGACGCTTCTCATTCC
DYSF_24	AGCTGGGAGCAGTTGTCAAT	GCAGCTTTGGCTCTATGTCC
DYSF_25	TTCATGTTGGGTTGTTGTGG	CAGTCCTGGGAGAGTTCAGC
DYSF_26	AATCACTTGAAAGGGTAGGGA	CAGTCCTGGGAGAGTTCAGC
DYSF_27	TCCTCAAAGACACCCAGGAC	ATTTGGCTGAGATCCCTCCT
DYSF_28	TTGGTTGGCATTCAACTCTG	CAGGTCTGCATCTGTGCCTA
DYSF_29	CTCCAGGAGGTGGTAGATGG	GATCTGTGGGTGTTCCCAGT
DYSF_30	GCTGTGGTTGGGAAATAGGA	CTGGATTTCAGAGGGAGCAG
DYSF_31	AAGTGGTCCAGTCTTGGTGC	CGAAAGCCAGATGTCTCCAT
DYSF_32	ATCTGCCATAACCAGCTTCG	AGGGACTTGTCTGCTGTGCT
DYSF_33	CTCACAGACACCAGCAGCTC	CAGCCCATAGCACTCTCTCC
DYSF_34	GAGGAAGAGTCCATGTGGGA	CCATGGTTTGCAGCCTCTAT
DYSF_35	GTTTATGGGTCGCTGCATCT	GCAGCTGAACTTGGCATGTA
DYSF_36	GCACTGGATGCATTACCTGA	GGGCTCTCCTTCCTGTCTCT
DYSF_37	CTTTCTGGCTCACAATGCAA	CAGACCTGCCTTACTCTGGC
DYSF_38	CTTTCTGGCTCACAATGCAA	GCTTCTGTTGACAGCCACTG
DYSF_39	GCCTAGACCTAGTGGCCAGA	GGGCTCCTTGTCATCAATGT
DYSF_40	GGAGAGCTTCCTGTGTGACC	AGGGTGACAACCTGGAACAG
DYSF_41	AGGTCAGGATTTGCCACAAC	CACAGAAACAGGGTTTCCCA
DYSF_42	AACCTGTGTCACTTGCATAATTAAA	GGGTCACCAGTGTAGGTACGA
DYSF_43	GAAGACATACCCAAGACTTGG	ACCTGGGACTCTGCCATGA
DYSF_44	CTTGAAGCCTTCCTGATGCT	CCTCTAGCTCTTGCTACAAACACA
DYSF_45	AATTCTCCCTCCATCCCATC	GTCCAGAGCTGAGGAGCAAG
DYSF_46	ACAGGCTGCTGTCCAAGTTT	GCATCTCAGACACACGGAGA
DYSF_47	CCTAGCAGGGAGGAGCTGTA	GCATCCTCATGGCTCACTTT
DYSF_48	AAAGTGAGCCATGAGGATGC	TCTTCAAAGCCAATCATCCA
DYSF_49	CTGAACGGTGCTCTTTGACA	CTTTAGAAGCCCTGGTGCTG
DYSF_50	TCTTAAGGCCTTCCCATCCT	AAGCAACTCCCAATCCTGTG
DYSF_51	TTTCAGCAGGAGACGGAACT	CTGCTCTCACAGATGAGCGT
DYSF_52	TAATTGAAGAGGTGGGTGGC	TGCTTTGCAGACATTGGTAAT
DYSF_53	GAAATGCTCATTGCTGCTGA	TCCAGCAAACACATTCCTGA
DYSF_54	GAGACCCGTGAGACACCAGT	CCAAGTGAAAGGAAACCCAA

DYSF_55	GCTCTGTTTCCAGAGTTGGC	AATAGGCCAAAGCCAGAGGT

The primer sequences in Table 6 are SEQ ID NOs: 373-534, respectively (forward primer, reverse primer, from top to bottom).

Table 7. Primer Pairs Used to Sequence the CAPN3 and DYSF Exons and Promoters.

Gene_Exon	Internal Primer A	Internal Primer B
CAPN3_1	TCTCAGATGACAGAATTACTCCAA	CAGAGCTGCTGCCAGGAT
CAPN3_2	CTGGCCAACATGGTGAAAC	GATGCATGGCAGAGTGCTAA
CAPN3_3	CCTGTTGATCATATTGTCAAGGAA	AGGGATTAGGGAGCCAGAGA
CAPN3_4	GCACCCAGTCCAGTTAGAGA	TTAGAGCTGTTGTTGCCTGG
CAPN3_5	TCTTGGGTGGGTCACTTAGC	TCCCTTGAGAAATTCCCAGTC
CAPN3_6	ATGGACAGCTTGGAAGGTCA	CTGGTTCTTGCACCCTCTTC
CAPN3_7	TGGTCAGGACAGAGCCTTCT	AAACTGTGCACCAACTGTGG
CAPN3_8	AGATGGCCAAGCCCTAAGTT	CTTCCACTCCTGGCCCTT
CAPN3_9	TCACCAGCCCATTTAAGGAG	CTGGAATAGAGTGTGTGGCG
CAPN3_10	TCAGAAGTGACAGCGTTTGC	CAAGCAGCATCTGCATTGTT
CAPN3_11	CTCCATCTGAATAAAGGTAGCG	CGCTCCACTGCCTCTCTAAT
CAPN3_12	ATACTTTCCCAGGGAGGACG	GAGTGTGCAAAGGCATGTGT
CAPN3_13	ATTTAAGCCTTGGGAGTCGG	GCCTGGAACATAGTAGGTGCTC
CAPN3_14	CTCTGTCCTTGGAAGATGCAC	GACCCTCTTCCATATTTCCCA
CAPN3_15	CCTTGCCATATGCAGTAAGAG	TAGGGCTGTTGTGAGGAAGG
CAPN3_16	AGGAGGGATGGAGTGGGTAT	CCTGCCAGTCCACTCCTAGA
CAPN3_17	CGCCATATCTCCTTTGGCT	GCACCTCAGCTATCAGGACC
CAPN3_18	CACACAAATCCACAAGCCCT	CACCCTGTATGTTGCCTTGG
CAPN3_19	AACACAGCCAGGTGGAATTT	CAGGCCTGAGAGAAGCACA
CAPN3_20	TGTTGGGTTGTAACTGCCCT	ATTCCTGCTCCCACCGTCT
CAPN3_21	TAGACCCTCCCTCCAAATCC	GCTGGTTGTTGAGGTGGAAT
CAPN3_22	GAGATGCGAAATGCAGTCAA	AGCACAAAGATGTGCAGGC
CAPN3_23	TGATAATCTCCAGTCTGCTCCA	GCAGTGGCTTACTGTTTCCTTT
CAPN3_24	CAGGACACATGCACTTGAGG	ACTTTCCTCCACATGGCAAA
CAPN3_Ep1	ACAGAGTGCTGTGTGTTGGG	GACACTGGAGCGAAATGTCA
CAPN3_Dp1	TTGCATGACCCATGACTACC	CTTCCCAACTCCCTGGTCAC
DYSF_1	GAGCCTTTCTCCTGTCCAAG	CTAGGTGCTCTCCAGGGTTG
DYSF_2	TTAAGGAGAGTCAGCCTGGG	CAAGAGAGTCCCTGAGCACC
DYSF_3	GGGTTGAAACTGAGTTGGGA	GGAAGCTCAGCTGTACCCAT
DYSF_4	TTCCCATGCCCAAGTATTTC	CCTCTGCCCTTCCCATCT
DYSF_5	GCCTAAGGTCACACAGCTCC	CACATTACTCCCTGCACCG
DYSF_6	GACTGCCCTCAAGTTTCAGC	AACTCCCTGTTTGGCATCTG

DYSF_7	CAGCCTGGCAGCTCTTCTAT	ATAGGGTGACAGGGATGTGG
DYSF_8	TCTGTGGGACTGGAGAAAGG	TTCTGTGACCCGTAGAGCCT
DYSF_9	TATGCCGTGTAGGGATTGTG	AGAGGGCTTGGCGTTGTTC
DYSF_10	CTCCCAAAGTGCTGGGATTA	GCTTGTCACCCAAATGACCT
DYSF_11	CAGCCTCTTACAGGCGTTTC	CAGAGGGATGTGCAATGAGA
DYSF_12	ACTGGAGATGTTCCTCGCAC	AGGACATTGGAATGGAGCTG
DYSF_13	AGCTGTTTGGGACTGGTGAC	CAGACCTGTCCACATTCGTG
DYSF_14	GTAGAAGGGCTGTGGCATTC	CGCCCTAAAGACTCCAAGAC
DYSF_15	CCCTGTGTCTTCTAGCTGTGC	CTGCCCTCAGAGATGATTCC
DYSF_16	GCGTCTGTAGAGATCCAGGC	GGCATATCCCACAATCCAAG
DYSF_17	CGGAACACACAGAGTGATGG	TCTAACTCGAGCATCAGCCC
DYSF_18	TTCTTTGCATCTCCAAGCCT	CATGGAAGGATCAGACTGGC
DYSF_19	CATCTGGGTGGCTTGTCATA	GAAGCAGGGCAAGTGTTGAT
DYSF_20	ATGCTGTTTCTTTCTTGGGC	AATGATCAGGATGGGTCAGG
DYSF_21	CACTAGGGAACACGGGTACG	TCTGTGTCCCACTGCACACT
DYSF_22	AGACTGGATGTATTTGGGCG	GCTGCTGCAGGGAGATTTAT
DYSF_23	AGATGGCTGTGTGTGTGGAG	TTCCTTCTGCAAATTGGTCC
DYSF_24	GCCACTCAAGCCAGACACT	TGATTCCGGCTCAAACCTAC
DYSF_25	GGAATGATGTAGCCTTTGCC	TTGGGTAGCTTGATCTTGCC
DYSF_26	GATACGGGTCAAGCTGTGGT	CAGTCCTGGGAGAGTTCAGC
DYSF_27	TCTCGGAGTGTCCCTAGGTC	GGCAAGCAATGAGAGGAGAC
DYSF_28	TACCTCCGGAGACTTCATGC	CTCCTGGGACCATCTCTGAA
DYSF_29	CCCTTCACTGGGCTATTTCA	ATCTTTGGGTATGCTGGGTG
DYSF_30	TTCCTGTGGCTGCAGAAAG	AGCAAGTGTTTCAGTGCCAA
DYSF_31	TTCCGTTCTGACTCATCTGG	GGGCCTTAAATGCCTGATCT
DYSF_32	TGTGGCTGTCCCATTGTCTA	TCAGCGAAGCCTGATCCTAC
DYSF_33	AGGACCCAGGCTCCATGT	GCATCTGTGCTAGCAATCCA
DYSF_34	GTCACCACAGGCTGCTCAC	AACCACGTCAGGAGATGACC
DYSF_35	TGGGTTGGACCTGTACCTTC	TCCTTCCATCTGGGATTCTG
DYSF_36	GCACTGACATCCATCACACC	TTGTCTGGGTGAAATCTGGC
DYSF_37	GGTGCTGGAATTGTGATCCT	GCAGATGTCAAAGTTGGGGT
DYSF_38	GAGGGAGGCCAACATCTACA	CTGAACCCTTCCAGTGAGGA
DYSF_39	TGAACAGGATGCATTTGGAA	CCTAAGGAAGGTCTCCACCC
DYSF_40	AGAGAGGGCAGGGAGACAAT	GGATTGAGTCTTGCCCAGAT
DYSF_41	CCAACCAAATGCTGAAACCT	GTTATCCCAGCCCACACTTG
DYSF_42	GTTCCTTTCTGGCTCCCTCT	AACACCATCCCATCACCAGT
DYSF_43	CACGAGAATAGCATGGGAAA	TACTGACACTGGCCTTCCCT
DYSF_44	TGTTTCTGATAAGGGCCTGG	GGAGCTTCTGTTGGGATCAA

DYSF_45	ACACTCAGGCCCAGTACAGC	TGTGATGAGCCAGGTTCTTG
DYSF_46	TGAGCCTCCATTTCTCCATC	CAGTGGCATCACAGGTCAGT
DYSF_47	AAGCCTGGAGCTAGTGGACA	CAGAGGAAGCCAGGACCTAA
DYSF_48	ATCTCTGAGAAGCCCACCCT	GAAGCCAAGAAGCAGACTGG
DYSF_49	AGAGCCAGAAGGTGACTTGC	CAACCCAAAGTTCAGTGCAG
DYSF_50	TGCACTGAACTTTGGGTTGA	AGACAGCAGTGGTGGTGACA
DYSF_51	TTGGGAGGATTAATGGAGCC	ACCTCTACTGACAGGCCCAC
DYSF_52	GATGGAATGGGAGACAATGG	GGGAGGAAAGAGGGAGAATG
DYSF_53	GCTATGATGCATGCAAATGTT	CTGCATCTTGAATTCGCTGA
DYSF_54	CAGCACCCAGAAGAGGAGG	GGACTAAGAGCCTCCAAGGG
DYSF_55	GTCCTCTCCCAGCCTCTTG	ACTGCTTCTCAGCTGCCTCT

The primer sequences in Table 7 are SEQ ID NOs: 535-696, respectively (internal primer A, internal primer B, from top to bottom).

Program Listing

The following is a program listing of an example of a Perl script for the analysis of primers for use in the disclosed method.

```
#!/usr/local/bin/perl

##########################################
#### Primer Prediction Utility
##########################################

use Getopt::Std;
use Bio::Seq;
use Bio::SeqIO;
use Bio::SeqI;
use Bio::SeqFeatureI;
use Bio::Tools::CodonTable;
use Getopt::Std;
use Cwd;
use Getopt::Std;
use Storable qw{dclone retrieve store};

#### Get Parameters
getopt('o::l::L::s::p');

### Error out if the required parameters are not passed
if (!$opt_o || !$opt_s || !$opt_l || !$opt_L) {
   die "Usage: single_primers.pl -o SEQOBJ.store -l Smallest -L Largest -s GenomicFlank
to grab
(-p 1 * for PCR primers, leave off if for sequencing primers)\n\n";
}

#### Get Bio::Seq Object
```

```
eval{  $in=Bio::SeqIO->new('-file' => "$filename",
                           '-format' => 'GenBank');
   };
$seqobj = retrieve "$opt_o";

#### Retrieve Exons for the Seqobj
(@exons) = &feature_array("exon");
if ($exons[0] == -1) {
   die "No Exons in $opt_o\n";
}
$exon_number = scalar(@exons);

#### Make a genomic file
&make_genomic;
print "There are $exon_number exons\n";

#### Process the exon info
$exonc = 0;
print "Processing Exon Info\n";
foreach (@exons) {
   $exonp++;
   $start = $_->start();
   $end = $_->end();
   print "START $start -> $end\n";

   $size = $end - $start;
   $flank = $end;
   $flank -= $start;

   ### calculate the distance for the exon from the end of the sequence segment
   ### and then extracts the segment of sequence with the exon centered in it
   if ($flank < $opt_s) {
         $flank = $opt_s - $flank;
         $flank /= 2;
         $flank = sprintf ("%.0f", $flank);
         $start -= $flank;
         $end += $flank;

   } else {
         $start -= 250; ## for sequence
         $end += 250; ## for sequencing
         $flank = 250;
   }

   $exoncoords{"$exonp"} = "$start,$end";
   $flank{"$exonp"} = $flank;
   $size{"$exonp"} = $size;
#    print "$exonp = $start,$end\n";
}
```

```
#### Now that we have exon info lets get the sequence
(@GENOMIC) = split(//,$seqobj->seq());

### if PCR Primers mask Repeat Elements (Repeats are marked in the seqobject)
if ($opt_p) {
   my $temp;
   (@Repeats) = &feature_array("misc_feature","note","RepeatMask");
   foreach $r (@Repeats) {
        $start = $r->start();
        $end = $r->end();
        $temp = $start;
        while ($temp <= $end) {
           $GENOMIC[$temp-1] = "N";
           $temp++;
        }
   }
}

#### Lowercase all exons
(@e2) = &feature_array('exon');
foreach $r (@e2) {
   $start = $r->start();
   $end = $r->end();
   $temp = $start;
   while ($temp <= $end) {
        $GENOMIC[$temp-1] =~ tr/[A-Z]/[a-z]/;
        $temp++;
   }
}
$total_g = scalar(@GENOMIC);
print "Total bases = $total_g\n";

#### now that i have the genomic i am going to extract the exon genomic ( minus 100 bases
for the sweet spot of sequencing)

print "Partitioning Exon Sequence\n";
foreach (sort keys %exoncoords) {

   ($start, $end) = split(/,/,$exoncoords{"$_"});
   $start -= 1; #want 100 bases not 99
   $end += 1;
    print "Coord = $_ $start,$end\n";
#   print "$start, $end\n";
   $glob_start{$_} = $start;
   $glob_end{$_} = $end;
   $basec = 0;
   foreach $agct (@GENOMIC) {
        $basec++;
        if ($basec == $start) {
           $base_on = 33;
```

```
        }
        if ($basec == $end) {
          $base_on = 87;
        }
        if ($base_on == 33) {
          if ($agct =~ "G" || $agct =~ "C") {
                $gc++;
          }
          $exonsequence{"$_"} .= $agct;
        }


    }
    $gc_content = $gc;
    $gc = "";

    ####### Mask Sequence Runs
    $exonsequence{"$_"} =~ s/GGGGGG/NNNNNN/g;
    $exonsequence{"$_"} =~ s/GGGGG/NNNNN/g;
    $exonsequence{"$_"} =~ s/GGGG/NNNN/g;
    $exonsequence{"$_"} =~ s/CCCCCC/NNNNNN/g;
    $exonsequence{"$_"} =~ s/CCCCC/NNNNN/g;
    $exonsequence{"$_"} =~ s/CCCC/NNNN/g;
    $exonsequence{"$_"} =~ s/TTTTTT/NNNNNN/g;
    $exonsequence{"$_"} =~ s/TTTTT/NNNNN/g;
    $exonsequence{"$_"} =~ s/TTTT/NNNN/g;
    $exonsequence{"$_"} =~ s/AAAAAA/NNNNNN/g;
    $exonsequence{"$_"} =~ s/AAAAA/NNNNN/g;
    $exonsequence{"$_"} =~ s/AAAA/NNNN/g;
}

### Create directories
if ($opt_p) {
   if (!-d "pcr_pr3") {
         `mkdir pcr_pr3`;
   }
   $dir = "pcr_pr3";
   $oli_file = "PCR_OLI";
} else {
   if (!-d "seq_pr3") {
         `mkdir seq_pr3`;
   }
   $dir = "seq_pr3";
   $oli_file = "SEQ_OLI";
}

#### Generate an error log
open(ERROR, ">$dir/error.log");
print "Printing Sequnece Info\n";
open(EXONFASTA, ">$dir/exons_seq_fasta");
```

```
open(DMDOLI, ">$oli_file");
foreach (sort keys %exoncoords) {

   ($start, $end) = split(/,/,$exoncoords{"$_"});

   $flank = $flank{"$_"};

   $target_start = $opt_s - $opt_l;
   $target_start /= 2;
   $target_start = sprintf("%.0f",$target_start);

   $target_size = $opt_l;
### Target size is the smallest acceptable product size
   $target_size = sprintf("%.0f",$target_size);

   open(EXONIND, ">$dir/EXON_$_\_FASTA");
   open(PR3TEMP, ">$dir/PR3.tmp");

   print EXONFASTA ">EXON_$_\n";
   print EXONIND ">EXON_$_\n";

   print PR3TEMP "PRIMER_SEQUENCE_ID=EXON_$_\n";

   $exonsequence{"$_"} =~ tr/[X]/[N]/;
## Some sequence has X's intead of NN's, primer 3
doesn't like X's

   print PR3TEMP "SEQUENCE=$exonsequence{$_}\n";

   (@exons) = split(//,$exonsequence{"$_"});
   $exon_seq_count = scalar(@exons);

   print PR3TEMP "TARGET=$target_start,$opt_l\n";
   print PR3TEMP "PRIMER_NUM_NS_ACCEPTED=0\n";
   print PR3TEMP "PRIMER_PRODUCT_SIZE_RANGE=$opt_l-$opt_L\n";

   print PR3TEMP "PRIMER_EXPLAIN_FLAG=1\n";
    print PR3TEMP "=\n";
   close PR3TEMP;

   print "Exon $_ has $exon_seq_count in its PCR Region\n";

   $basec = 0;
   $nl = 60;
   foreach $e (@exons) {
         $basec++;
         print EXONFASTA "$e";
         print EXONIND "$e";
         if ($basec == $nl) {
            print EXONFASTA "\n";
```

```
            print EXONIND "\n";
            $nl += 60;
        }
    }
    print "Picking Primers for $_\n";

    @primer3 = `primer3 < $dir/PR3.tmp > $dir/EXON_$_\_PR3`;
### PRIMER3 Prediction program

    close EXONIND;
    print EXONFASTA "\n";

    ### Lets Process the PR3 Output
    chomp($left_pcr_pos = `grep "PRIMER_LEFT=" $dir/EXON_$_\_PR3`);
    chomp($left_pcr = `grep "PRIMER_LEFT_SEQUENCE=" $dir/EXON_$_\_PR3`);
    chomp($left_pcr_tm = `grep "PRIMER_LEFT_TM=" $dir/EXON_$_\_PR3`);
    ($label, $left_pcr) = split(/=/,$left_pcr);
    ($labe,$left_pcr_tm) = split(/=/,$left_pcr_tm);
    chomp($right_pcr_pos = `grep "PRIMER_RIGHT=" $dir/EXON_$_\_PR3`);
    chomp($right_pcr = `grep "PRIMER_RIGHT_SEQUENCE=" $dir/EXON_$_\_PR3`);
    chomp($right_pcr_tm = `grep "PRIMER_RIGHT_TM=" $dir/EXON_$_\_PR3`);
    ($label, $right_pcr) = split(/=/,$right_pcr);
    ($label,$right_pcr_tm) = split(/=/,$right_pcr_tm);
    undef($lglobal_start);
    undef($lglobal_end);
    undef($rglobal_start);
    undef($rglobal_end);
    if ($left_pcr_pos =~ /\d+,\d+/) {
        ($j,$pos) = split(/=/,$left_pcr_pos);
        ($st,$len) = split(/,/,$pos);
        $lglobal_start = $glob_start{$_} + $st + 1;
        $lglobal_end = $lglobal_start + $len;

        ($j,$pos) = split(/=/,$right_pcr_pos);
        ($st,$len) = split(/,/,$pos);
        $rglobal_start = $glob_start{$_} + $st - 1;
        $rglobal_end = $rglobal_start - $len;

        open(OLI, ">$dir/EXON_$_\_OLI");
        print OLI ">EXON_$_\_LEFT TM:$left_pcr_tm\n";
        print OLI "$left_pcr\n";
        print OLI ">EXON_$_\_RIGHT TM:$right_pcr_tm\n";
        print OLI "$right_pcr\n";
        close OLI;
    }

    print DMDOLI ">EXON_$_\_LEFT TM:$left_pcr_tm START:$lglobal_start END:
$lglobal_end\n";
    print DMDOLI "$left_pcr\n";
    print DMDOLI ">EXON_$_\_RIGHT TM:$right_pcr_tm START:$rglobal_start
```

```
END:$rglobal_end\n";
  print DMDOLI "$right_pcr\n";

  if (! $left_pcr || ! $right_pcr) {
       print ERROR "EXON_$_ NO PRIMER\n";
  }
}
close EXONFASTA;
close ERROR;
close DMDOLI;

### Masked Sequence Subroutine
sub make_masked {
   (@genomic) = split(//,$seqobj->seq());
   (@Repeats) = &feature_array("misc_feature","note","RepeatMask");
   foreach $r (@Repeats) {
        $start = $r->start();
        $end = $r->end();
#       print "$start -> $end\n";
        $temp = $start;
        while ($temp <= $end) {
           $genomic[$temp-1] = "N";
           $temp++;
        }

   }
#   die;
   open(MASK,">$opt_o.masked");
   print MASK ">$opt_o\_masked\n";
   $lb = 50;
   $c = 0;
   foreach $g (@genomic) {
        print MASK "$g";
        $c++;
        if ($c == $lb) {
           print MASK "\n";
           $c = 0;
        }
   }
   close MASK;
#   die;
}


### Genomic Output Subroutine
sub make_genomic {
   $seq = $seqobj->seq();
   $genomic_query = "$opt_o.genomic";
   open(GENOMIC,">$opt_o.genomic");
   print GENOMIC ">TEMP\n$seq\n";
```

```
  close GENOMIC;
}

### Feature retrieval subroutine
sub feature_array {
  undef(@returns);
  ($tag) = $_[0];
  ($subtag) = $_[1];
  ($subvalue) = $_[2];
  @all = $seqobj->all_SeqFeatures();
  foreach (@all) {
      if ($_->primary_tag =~ /$tag/) {
        if ($subtag && $subvalue) {
            eval{
              ($cvalue) = $_->each_tag_value("$subtag");
              if ($cvalue =~ /$subvalue/) {
                  push(@returns,$_);
              }

            };
        } else {
            push(@returns,$_);
        }
      }
  }
  if ($returns[0]) {
      return(@returns);
  } else {
      return(-1);

  }

}
```